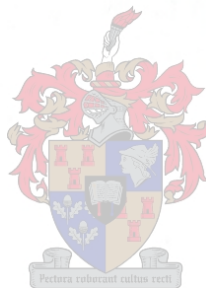


Complexity and the Self

Tanya de Villiers



Thesis presented in partial fulfilment of the requirements for the degree
of Master of Arts at the University of Stellenbosch.

Thesis supervisor: Professor F. P. Cilliers

December 2002

Acknowledgements

I want to thank my supervisor, Prof. Paul Cilliers, the Department of Philosophy at Stellenbosch University, my family, and a special thank you to my friends for their various contributions of time, expertise, encouragement and support in the writing of this work. A special thank you to Johan Hugo for his meticulous and invaluable proofreading of the text.

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at are those of the author and are not necessarily to be attributed to the National Research Foundation.

Declaration

I, the undersigned, hereby declare that the work contained in this thesis is my own original work and has not previously in its entirety or part been submitted at any other university for a degree.

Abstract

In this thesis it is argued that the age-old philosophical “Problem of the Self” can benefit by being approached from the perspective of a relatively recent science, namely that of Complexity Theory. With this in mind the conceptual features of this theory is highlighted and summarised. Furthermore, the argument is made that the predominantly dualistic approach to the self that is characteristic of the Western Philosophical tradition serves to hinder, rather than edify, our understanding of the phenomenon. The benefits posed by approaching the self as an emergent property of a complex system is elaborated upon, principally with the help of work done by Sigmund Freud, Richard Dawkins, Daniel Dennett, and Paul Cilliers. The aim is to develop a materialistic conception of the self that is plausible in terms of current empirical information and resists the temptation see the self as one or other metaphysical entity within the brain, without “reducing” the self to a crude materialism. The final chapter attempts to formulate a possible foil against the accusation of crude materialism by emphasising that the self is part of a greater system that includes the mental apparatus and its environment (conceived as culture). In accordance with Dawkins’s theory the medium of interaction in this system is conceived of as memes and the self is then conceived of as a meme-complex, with culture as a medium for meme-transference. The conclusion drawn from this is that the self should be studied through narrative, which provides an approach to the self that is material without being crudely physicalistic.

Abstrak

In hierdie tesis word daar aangevoer dat die relatiewe jong wetenskap van Kompleksiteitsteorie 'n nuttige bydra kan lewer tot die eeue-oue filosofiese "Probleem van die Self". Met die oog hierop word die konseptuele kenmerke van hierdie teorie na vore gebring en opgesom. Die argument word gemaak dat die meerendeels dualistiese benadering van die Westerse filosofiese tradisie tot die self ons verstaan van die fenomeen belemmer eerder as om dit te bemiddel. Die voordele van dié nuwe benadering, wat die self sien as 'n ontluikende (*emergent*) eienskap van 'n komplekses sisteem, word bespreek met verwysing na veral die werke van Sigmund Freud, Richard Dawkins, Daniel Dennett en Paul Cilliers. Daar word beoog om 'n verstaan van die self te ontwikkel wat kontemporêre empiriese insigte in ag neem en wat die versoeking weerstaan om ongeoorloofde metafisiese eienskappe aan die self toe te ken. Terselfdetyd word daar gepoog om geensins die uniekheid van die self te "redukeer" na 'n kru materialisme nie. In die finale hoofstuk word daar gepoog om 'n teenargument vir die voorsiene beswaar van kru materialisme te ontwikkel. Dit word gedoen deur te benadruk dat die self gesien word as deel van 'n groter, komplekse sisteem, wat die masjienerie van denke en die omgewing (wat as kultuur gekonseptualiseer word) insluit. Insgelyks, in die teorie van Dawkins word die medium van interaksie in hierdie sisteem gesien as "memes", waar die self dan 'n meme-kompleks vorm, en kultuur die medium van meme-oordrag is. Daar word tot die konklusie gekom dat die self op 'n narratiewe manier bestudeer behoort te word, wat dan 'n benadering tot die self voorsien wat materialisties is, sonder om kru fisikalisties te wees.

*Opgedra aan Dr. H. F. Kotzé en Prof. F. P. Cilliers, sonder wie hierdie werk
nie moontlik sou wees nie.*

Contents

| | |
|-----------------|----------|
| Preface, | 7 |
|-----------------|----------|

| | |
|-----------------------------------|-----------|
| 1. A Primer on Complexity, | 10 |
|-----------------------------------|-----------|

| | |
|---|----|
| 1. Introduction, | 10 |
| 2. What is a Complex System?, | 15 |
| 3. Non-linear interactions, | 18 |
| 4. Open Systems, | 19 |
| 5. Emergence, | 22 |
| 6. Framing and Modelling, | 26 |
| 7. Self-organisation, | 29 |
| 8. Feedback, the environment, and adaptability, | 31 |
| 9. Information and enabling constraints, | 37 |
| 10. Conclusion, | 40 |

2. Landmarks in the History of the Self: The Legacy of the Enlightenment, 44

1. Soul, self, consciousness, ego, subject, mind, identity, self-concept, personality – semantic nitpicking or rigorous distinction? 44
2. Introducing the self, 46
3. An ancient self, 48
4. Some modern metamorphoses of self, 56
 - 4.1 Intimations of Rationality, 57
 - 4.2 Descartes criticised, 67
 - 4.3 Intimations of Empiricism, 72
 - 4.4 The self becomes transcendental, 81
5. Fast forward to the twentieth century, 91

3. A Neurological Basis for the “Higher Functions” of the Brain, 93

1. Introduction, 93
2. The self as grammatical error, 95
3. Freud’s “new mind”, 99
4. The Project, 101
5. Freud’s differentiated neurons become conscious, 106
6. In the realm of the unconscious, 112
7. A note on the “Mystic Writing-Pad”, 119
8. Post-Freudian theory, 124
9. Dennett’s materialism and the key to a demystified “mind”, 126

4. The Complex Self, 135

1. Introduction, 135
2. Some of the complexities of brain structure in a nutshell, 138
3. Where to from here? 148
4. Dennett's material self 149
5. The self as a complex system 151
 - 5.1 Cartesian conceptions of the self disqualified, 152
 - 5.2 Freud lays the foundation, 157
 - 5.3 The material basis of the self revisited, 160
 - 5.4 Dennett's self revisited, 163
 - 5.5 The material and complex self, 166

5. On the Subject of the Human Environment, 175

1. Introduction, 175
2. Nagel's mysterianism and bats, 177
3. *Culture* red in tooth and claw? 181
4. Plato's view on "art", 187
5. Popper's world 3, 191
6. The complex self and the changing role of a uniquely human environment, 196

Preface

Questions concerning the self are some of the oldest questions in philosophy. And, as seems to be the case with most philosophical questions, the more one tries to study or define this elusive “I”, the more intangible it seems to become. In surveying the literature on this subject, it seems that one can never quite be satisfied that what one is reading is getting to the heart of the problem. One reason for this is that there is no real consistency in the way that concepts pertaining to the self, such as “the I”, “ego”, “personality”, “consciousness”, “cogito”, etc., are applied. All of these concepts crop up somewhere along the line, seemingly denoting the same phenomenon. At the same time, depending on whom one is reading, there could also be subtle distinctions between the denotations of these different terms – subtle distinctions with critically important implications.

This work will not be an extended exercise in conceptual clarification, however. This is partly because we would contend that such an exercise would be, to a significant extent, ineffectual. The reason for this ineffectuality may not so much lie in a shortcoming in the concepts that are used, or in the way that they are generally applied, but in the phenomenon itself – in these elusive things that we lump together as aspects of the “self”. Bearing this in mind, this work will focus mainly on what kinds of things have been relegated to the realm of self, consciousness, even mind, and to look at the way that certain presuppositions concerning the kind of phenomena that belong to this realm have influenced resultant theories.

As we shall see, the question as to what human beings are, in essence, is as old as philosophy itself. What seems to be an especially

contentious issue in the Western philosophical tradition is the materiality or non-materiality of this essence. Is the self, the soul, consciousness, the truly human part of “us” simply an extension of the material body – a useful evolutionary tool that optimises our chances for survival? Or is it some kind of essence – divine, spiritual or reasonable – that animates the material body and endows it with characteristic that elevates it above the grossly material?

Of course, many factors come into play in the case of such questions, ranging from theoretical, political and religious contexts and agendas to the availability of information and established models with which to work. Central to *this* work will be the possibility of developing a contemporary model of the self that is based on a contemporary science. We hope to explore how bringing a new theoretical angle into the debate may highlight possible alternative perspectives, and perhaps lead to fresh, interesting, and even useful new insights on this ancient topic. A contemporary discipline that could prove useful in this regard is that of complexity theory.

Keeping this in mind, the first chapter will consist of a brief overview of the main tenets of complexity theory. Readers already familiar with the subject might want to proceed to the following chapter.

The second chapter will briefly examine the tradition in which current conceptions of the self were born. Inevitably such an enquiry would need to address Descartes’ dualism, which can safely be said to be the defining contribution to the debate in the Western philosophical tradition. Most of the genealogy of current conceptions of the self can be traced back to Descartes and his (relative) contemporaries, with the rationalist/empiricist debate. For this reason Hume and Kant will also be discussed in some detail.

While the rationalist/empiricist debate over the subsequent centuries can with some justification be described as a family feud, the late nineteenth and early twentieth centuries would bring a radical turning point in traditional conceptions of the phenomena associated with consciousness and the self. Whereas phenomena to do with the mental were generally associated with the immaterial, and where the self had always been associated with the conscious, a new possibility would enter the scene. Sigmund Freud would open up a whole new realm of possibilities with his postulation of the existence of the unconscious. Chapter 3 consists of a summary of Freud’s

view, and will also explore the impact that not only Freud, but also advances in the neurological sciences and psychology had on theories of the self. We will primarily be concentrating on the work of Daniel Dennett.

The fourth chapter is an attempt to consolidate all of these themes and to construct a contemporary model of the self, based on complexity theory. The feasibility and possible advantages of such a model are discussed.

Finally, chapter 5 will highlight some of the possible implications that a theory of the self, based on complexity, might have. One such implication would be the possible revision of both the concept of the human environment and the role that is usually ascribed to this environment as it pertains to the self. As can be imagined, such a project would have vast possibilities and far-reaching implications. So much so that it would range far beyond the scope of this work, and the discussion will be constrained to some of the most pertinent, and interesting, implications. Other possible implications that are raised will be mentioned, with the hope that they will be subjected to further investigation elsewhere.

Chapter One

A Primer on Complexity

Gone is the image of the clockwork universe

Auyang (1998)

One should be leery of these possibilities in principle. It is also possible, in principle, to build a stainless-steel ladder to the moon and to write out, in alphabetical order, all intelligible English conversations consisting of less than a thousand words. But neither of these two are remotely possible in fact, and sometimes an impossibility in fact is theoretically more interesting than a possibility in principle...

Dennett (1991:4)

1. Introduction

Since the argument hinges around the thesis that aspects of complexity theory can offer us a viable and useful model of the self, a brief overview of complexity theory may be useful. Theories of complexity, rooted primarily in the physical sciences spawn literature with a high degree of mathematical content, which is likely to be inaccessible to the lay reader. This, coupled with the vast amount of pop-science books that have appropriated some of the ideas of complexity, may create the impression that it is impossible to give a thoroughly qualitative reading of the theory, without veering off into the esoteric or the mundane. We would contend that it is indeed possible to give a general reading that touches upon the central aspects and ideas of the field, without losing too much content. Hence, a philosophical account of various theories of complexity might busy itself with the underlying assumptions and conceptual structures of such theories and explore their applicability to other spheres of thought.¹ Naturally, such an overview would lack the depth and intricacy of a mathematical theoretical account. Nevertheless it could prove useful to introduce the field to an audience with little mathematical knowledge, but who might still find the ideas

¹ I agree with Auyang (1998:X) that philosophy as a discipline seeks "general patterns of thought."

encompassed in it both interesting and useful with regard to their own fields of speciality.

The idea that science and indeed all theory is context-bound and reflects or encompasses a prevailing and general theoretical paradigm is nothing new to contemporary philosophy (Kuhn: 1970, Cilliers1998:1-2; 87-88). There is a long standing tradition in Western philosophy that argues that our experience and knowledge is structured by a framework of categories or general concepts; a categorical framework being the most basic of presuppositions about the intelligible world, and our relation to it (Auyang1998:XI). These categorical frameworks permeate our thinking and find their way into all our theories of the world. Classical science operated against a background of reversibility and timelessness (Prigogine 1984:7) and many complexity theorists would argue that this presupposition has led to a crude model of the universe which does not provide for the complexities of actual existing phenomena (Auyang 1998:X; Cilliers1998: 9-10).

Central to classical Western science is the analytical method. In order to grasp a phenomenon it is divided into manageable units, which can be studied and strung together again, to form a picture of the whole. The main criticism regarding this approach from some contemporary theoretical circles is its inability to allow for the relationships between the artificially separated components that to a large extent constitute the system's make up in the first place (Kauffman 1993:vii). Cilliers (1998:2) goes so far as to assert that in cutting up the analysed system this method destroys exactly that which it wants to explain. The person whom many would regard as a pioneer of complexity theory as we know it today, Ilya Prigogine, uses the metaphor of reducing a building to a pile of bricks, in the sense of not being able to see the building for the bricks, to describe this reductionistic method (1984 :7).²

In discussing the rise of complexity sciences it might prove useful to give a brief overview of the background against which this (very broad and

² This metaphor is somewhat misleading in that it does not capture the idea of relational dynamism, interaction, and flux, which complexity theory places so much emphasis on.

very diverse) discipline originated. The discussion will follow the lead of Prigogine and discuss the evolution of the ideas encompassed in this theory with reference to the developments in the Western physical sciences, with emphasis on theories that deal with macroscopic phenomena. He makes the telling observation that most of the scientific disciplines exhibit more or less the same characteristic changes over roughly the same time frame (1984:10).

In the seventeenth century Newton's physics allowed for what Prigogine dubs "a scientific revolution" (1984:28), which ushered in a new paradigm in what was then called *natural philosophy*. With the help of Newton's laws of motion scientists were now able to predict trajectories and formulate complete descriptions of dynamic systems. His law of gravity could be applied equally to explain the motion of planets and of atoms, and to explain why bodies fall back down to earth. Newton was hailed as a scientific hero, having discovered the basic laws of the universe, where every physical body has a mass and acts as a source of the forces that are a prerequisite for interaction. The basic characteristics of the resulting trajectories are lawfulness, determinism and reversibility (Prigogine 1984: 59-60). In order to calculate a trajectory, all that is needed is an empirical definition of a single instantaneous state of the system. This state can be thought of as an "initial state" and from such a state a subsequent series of states that the system will pass through in time (its trajectory) can be predicted. The implication being that the whole past and the future of a system can be deduced from a single given state; all states in the trajectory are equivalent and can be used to calculate all other states. An important prerequisite for this possibility is the reversibility of the trajectory of the dynamic system (61). A reversible trajectory is time independent, and the states of the system are, at least in principle, reversible.

Newton's theories were timeless and reversible, thus universally applicable. The resultant view was that reversible processes are the norm, while processes which are irreversible and time dependent, are the exceptions, anomalies. The only difference between simple and complex systems was that complex systems needed a complex description, which may or not be beyond current scientific capabilities – progress in science would

undoubtedly eventually lead to complete descriptions of even complex phenomena. All systems, in principle lent themselves to scientific description.

In the nineteenth century the new sciences of heat were to present a challenge to Newton's science of gravity (12). Fourier, with his law of the propagation of heat, presented, for the first time, a quantitative description of an irreversible process, something inconceivable in classical dynamics. Thermodynamics and its second law,³ reintroduced time into physics, providing what theorist like Prigogine regard as a conceptual revolution in the physical sciences (1984: xxviii). Newton's theories now had to contend with processes dependent on the direction of time, and which were for this reason, irreversible. Irreversible processes include processes such as chemical reactions, heat conduction and diffusion (Prigogine 1980:5).

The result, in Prigogine's view, is a scientific heritage that generates two questions, which it cannot contend with (1984:xxix). The first is the relation between order and disorder. The second law of thermodynamics calls for an increase in entropy, where, in the course of time a system would maximise its entropy and where disorder seems to be the natural state for a dynamical system, leading to the "heat-death of the universe". Yet at the same time we are confronted with the phenomena of biological and social evolution, where order and complexity seem to be on the increase. Structure arising from the disorder as Prigogine (*ibid.*) puts it. Secondly, both classical and quantum physics view reversible processes as the norm. This would seem to disallow evolution, where structures evolve towards greater complexity and where the reversibility of this process seems highly

³ According to the first law of thermodynamics the total amount of energy in the universe is conserved. The second law states that, with time, entropy (the disorderly arrangement of energy) will increase until it reaches its maximum value, which would result in a state of thermodynamic equilibrium (Prigogine 1980:5; Juarrero 1999:104) (Refer to section 3 for a more detailed discussion of the second law of thermodynamics and its implications.) The increase of entropy furnished classical thermodynamics with a criterion for differentiating between past, present and future. With the increase in entropy being irreversible we gain a measure for temporality, with a state of less entropy preceding a state of greater entropy. Thermodynamics and its laws only apply to closed systems, however, which are isolated from their environment and consequently do not exchange matter and energy with it. Hence, trajectories might not be reversible, but they are certainly predictable: all processes are heading towards equilibrium (Juarrero 1999:105). Classical mechanics and thermodynamics seem to be in accord about the deterministic nature of the universe.

improbable. A further difficulty, mainly with regard to classical biology, raised by Von Bertalanffy (1973:44), is the Darwinian view of evolution where organisms appear as the random products of chance – haphazard products of the undirected mutations and random selections. Similarly the mental world was seen as a curious epiphenomenon of events in the material world. Arbitrary and peripheral phenomena do not make sense in a universe based on lawfulness, determinism and reversibility. With the reintroduction of time into physics it was recognised that irreversibility, far from being an aberration, is an essential aspect of the natural world and lies at the origin of most instances of self-organisation. Reversibility and determinism only applied to limited and simple cases in the world or to theoretical abstractions (Prigogine 1984:8).

In addition to the objection that classical science cannot account for time, Newton's laws were discovered to hold only for two-bodied systems. As soon as further elements were added to the system and the interaction between the various bodies came into play, it became necessary to simplify the system, which complies with the theoretical approach of classical dynamics: the analytic method. If a given system is too complex to be grasped as a whole, it is divided into its smaller constituents, which are examined separately and then put together again. Presumably the characteristics of the system will be a compilation or a superposition of the characteristics of its parts. As said earlier, many contemporary theorists would question this assumption, given the realisation that elements in a system necessarily *interact* with one another, and that many of the properties of a system arise as a result of these interactions. The individual elements of a system in isolation cannot display the characteristics of elements in interaction. The further realisation that systems usually comprise of much more than two elements, and that additional elements complicate the dynamics of a system, gave rise to a new approach to the systems theories.

2. What is a complex system?

The field of the complexity sciences is vast and its theories find their application in very divergent disciplines. Consequently the terminology used differs depending on the theorist and on the scientific field in question.⁴ Here we have elected to speak about “complex systems”, but there are many terms that approximately denote the same phenomenon, such as: many-bodied systems (Auyang), complex adaptive systems (Gell-Mann), dissipative structures (Prigogine), etc. And, seeing that theories of complexity study complex systems, a definition of what a complex system is seems to be in order.

Defining a complex system is not a simple matter. As we will argue, complex systems, in principle, do not lend themselves to definition. A definition entails an abstraction that is unable to allow for the actual, contingent characteristics of individual systems. Furthermore, all complex systems do not display the same sets of properties. What we can do, however, is give a very general, very sparse description, which can serve to orientate us with regard to our subject matter, but which, by no stretch of the imagination encompasses all complex systems. Von Bertalanffy’s definition of a system gets us off the mark. He defines a system as “a set of elements standing in interconnection” (quoted by Wuketits 1998:318). Von Bertalanffy also designates as “systems problems” the problem of the interrelations of a great number of variables (1973:xix).

This definition does not seem to have to much to it, but it does point to two of the most important considerations of the sciences of complexity: systems that consist of many elements, and the interactions between these elements. Complexity theory takes note of the fact that the composition of a system is not merely the result of the aggregation of its individual components, but is generated by means of the interaction between components. These individual components do not operate in terms of a predetermined *telos* or goal, but act in terms of their own properties and

⁴ Refer to Juarrero (1999) pp. 109-117 for a brief survey of some of the largely non-uniform terminology used in the field.

purposes. Furthermore, characteristics that appear complex and disjointed at a relatively small scale generally prove to form recognisable and stable patterns at a larger scale.

Both Cilliers (1998:ix) and Kauffman (1995:18-19) make the point that in discussing complexity it is necessary to engage with a specific complex system, mainly because complexity arises from the specific interactions of specific components. The whole point of a complex approach to systems is to try and avoid abstractions and generalisations, which may not be sensitive to the particularities and contingencies of a specific system, as much as possible. Kauffman further asserts that though we are unable to predict detail (in principle), we are able to predict "kinds of things". He believes that such theories serve to explain what he calls "generic properties" of dynamic systems (Kauffman 1995: 17).

Having many components does not necessarily make a system complex. Cilliers (1998:3) makes the distinction between complex and complicated. Systems can have large numbers of components, which can be analysed accurately - the individual parts do aggregate into the whole system. Things like aeroplanes, snowflakes and the Mandelbrot set⁵ can be considered to be merely complicated.⁶ Other systems are comprised of components with such intricate interactions, that only certain aspects of them can be analysed at a time. These analyses would more often than not cause some sort of distortion of the system. These systems are complex. The distinction between a complicated and complex system can often be a function of the distance that the observer takes from the system (1998:3).⁷ Gell-Mann calls this "coarse-graining" by which he means that we need to

⁵ The Mandelbrot Set is often presented as an example of complexity arising from simple rules as the result of non-linear interaction. As Peak and Frome would have it: "The Mandelbrot set is the prototypical fractal: so easy to generate, yet so complex in structure. It forces us to question the essence of our understanding of simplicity and complexity. Diversity of forms set at all levels of magnification. The slightest bit of non-linearity can lead to such complexity" (Peak and Frome 1994:243). This approach overlooks both adaptability and interaction with the environment - abilities which characterise complex systems.

⁶ Cilliers (1998) does allow for the ambiguity that new technology can create. Powerful new computers would generally not be considered to be "living" in the traditional sense, yet there is an argument to be made out for their complexity (3).

⁷ Cilliers (1989) calls this activity "framing" (32-46).

specify the level of detail at which the system will be described (1994:29).⁸ It is important to stress that this does not mean that complexity is merely a function of human observation and human description. Because the complexity of a system is a result of the interaction between the different components of the system, it occurs at the level of the system itself (Cilliers 1998:2-3), and it would be perfectly plausible to speak of a complex system, independent of a human observer. It seems to make as much sense to say that there are components in the world that interact in a complex manner, as it does to say that there are components in the world (the objections of Cartesian solipsism aside).

There are many levels of complexity. Important in this regard is the realisation that complexity remains at the level of the system. When we talk about levels of complexity we do not mean some meta-level description “above” the system nor an underlying source “below” the system. Where our influence does make itself felt is at our choice of the distance that we will take from the system (due to interest, technical constraints, available knowledge etc.). Complex systems exhibit many characteristics, many structures at various scales and undergo various processes at various rates and, most importantly, they exhibit the ability to acquire information about their environment and to change and adapt (Auyang1998:13; Gell-Mann1994:17). Complex systems also interact upon and change their environment. In the light of this ability complex systems are often characterised as living systems. This does not imply that all complex systems are biologically alive - language, social structures, the economy, organisms and the ecosystem are all complex.

A complex system has more possibilities than can be actualised (Cilliers 1998:2). The number of components that can be considered to be a part of a complex system and the number of possible interactions between these different components make for a vast amount of possible configurations

⁸ It is useful to note Auyang’s (1998) assertion that coarse-graining filters out insignificant details and serves to bring emergent properties into relief (4). These details are of course insignificant only in relation to a specific description and may gain in importance as the focus of our study shifts.

and possible future states for the system. As the system increases in size the number of possibilities increases accordingly. The difference between the enormity of the possibilities and the scarcity of actualisations underpins concepts like probability, contingency, temporal irreversibility and uncertainty (Auyang 1998:18). In attempting to account for these occurrences theoretically, the complexity sciences distinguish themselves from the modern scientific world-view as well as thermodynamics and evolutionary theory.

3. Non-linear interactions

Dynamic systems as characterised in Newtonian physics are linear systems. Crudely simplified this means that the trajectory of a system (its successive states through time) can be analysed as the sum-total of its normal modes. Auyang (1998:178) explains this by way of the vibration of a violin string. The sound that the string produces can be seen as a succession of its normal modes, strung together to form a harmonic motion. Analogously the behaviour of a linear system is the result of superposition:⁹ an aggregation of its parts. Thus a linear response to a disturbance is proportional to the magnitude of the disturbance. To cast this in complex vernacular: small causes have small effects. This results in a stable, and predictable system, where its final state is determined by its initial conditions. As seen in the previous section, such systems lend themselves to prediction, because a particular state in the system, coupled with its basic laws enables theorists to project past and future states with accuracy. The trajectory of the system is stable, predictable and time independent (it can, in principle, be reversed). In a low-energy equilibrium system the system moves to equilibrium and no additional energy is needed to maintain it (Kauffman 1995:20). Linear systems are also closely related to the principle of symmetry, where linear relationships are symmetrical and hence give rise to simple systems with transparent structures (Cilliers 1998:120), and when dynamic processes are

⁹ Auyang defines the principle of superposition as a combination of solutions yielding another solution (1998:234). Hence, when a difficult problem is encountered it can be broken into simpler problems; the solutions of the simpler problems can then be superposed on to the original problem and would presumably provide a solution.

reversible. As stated above, thermodynamics and its second law re-introduces time into the systems equation and shows linear, stable systems to be the exception rather than the norm. The important difference is that the dynamic process of a thermodynamic system is irreversible.

The second law of thermodynamics states that in a closed system, a certain quantity, entropy, must increase to a maximum, where the system comes to rest in a state of equilibrium (Von Bertalanffy 1973:38). Von Bertalanffy characterises entropy as a measure of probability. A system tends to the state of most probable distribution, which is a state of complete disorder. Indeed, Kauffman (1995:9) defines entropy as a measure of disorder.

This point can be illustrated with the classical example of a dark blue droplet of ink, which is dropped into a jar of still water. Inevitably the ink will diffuse through the water and tint it a light blue. The most probable state of the system (water + dark ink) is that the ink would spread evenly throughout the jar of water. In the words of Kauffman: "Left to its own devices the system will visit all possible microscopic configurations equally often" (9). This is an irreversible process - the likelihood that the ink molecules would reverse the process and reassemble into an inkblot being minimal. As maximum disorder seems to be the most probable state of distribution of a system, in order for order to be maintained some work has to be done on the system (Kauffman 1995:10).¹⁰

4. Open systems

A useful distinction to make when trying to come to grips with concepts like that of reversibility and irreversibility or linearity and non-linearity is that of closed and open systems. The distinction between closed and open systems rests with the system's relation (or lack of it) with the environment. Reversible processes are generally restricted to closed systems: systems that are

¹⁰ Gregory Bateson (1972:3-8) uses the simple analogy of a little girl's room that just cannot seem to keep itself tidy; it naturally tends to disorder and it takes effort on the part of the girl to keep things tidy.

isolated and do not interact with their environments. Open systems, on the other hand, continuously exchange energy and matter with their environment and exist by virtue of *their interaction with their environment*. Closed systems move to a state of equilibrium. Equilibrium structures are inert at a global level, and once formed, they can be isolated and maintained indefinitely. The final state of a closed system will be determined by its initial conditions, whereas, in open systems, the final state is in essence unpredictable. Whereas a closed system is an isolated system and exchanges no matter or energy with its environment, an open system can, by definition, not exist in isolation; it exists in its exchange of matter and energy (or information) with the environment. In this sense biological cells and cities are open systems in that they “feed” on matter and energy coming from outside of the system (Prigogine 1984:127). This is not a one-way process though, open systems contribute to their environment; they make up part of a greater system.

All living organisms are essentially open systems (Von Bertalanffy 1973:31). It is still possible for a complex system to have a structure - a steady-state - which is maintained at a distance from equilibrium and where the system is capable of doing work. In spite of consisting of continuous irreversible processes, the constant import and export of matter, the system stays constant in its composition (more about this in our discussion on self-organisation).

It is important to emphasise that the second law of thermodynamics applies only to *closed* systems, which undergo irreversible processes. Only irreversible processes contribute to entropy production (Prigogine 1980:5). Examples of irreversible processes include: chemical reactions, heat conduction and diffusion (5). The irreversible processes of the second law of thermodynamics leads to a kind of one-sidedness of time, the increase in entropy is associated with a positive time direction.¹¹ Introducing directionality into closed systems does not, in itself, translate into a radical overhaul of our understanding of open systems and the role that time and the environment play in their structures and their characteristics. As we have seen in the

¹¹ Refer to footnote 3.

previous section, many contemporary theorists believe open systems to be the norm and in the light of the characteristics of open systems it becomes necessary to revise the laws of classical science and even the second law of thermodynamics, which pertain to closed systems in a state of equilibrium.

Prigogine insists that the second law can be extended to include open systems, which do exchange energy and matter with their environment (i.e. living/complex systems). He states that there is an essential difference between the laws for systems at equilibrium and for non-equilibrium systems: where laws for equilibrium systems are universal, laws for non-equilibrium systems become very specific (1980:93). While in isolated (closed systems) the law only allows for the fact that the system will increase in entropy and come to rest in a state of maximum disorder, applied to non-equilibrium open systems, Prigogine speculates, it may account for a new type of structure, which he calls a *dissipative structure*. Furthermore, it might account for the coherence and organisation found in the non-equilibrium world around us. Consequently it may explain why living systems, which are by definition not closed, and maintain themselves through a continuous inflow and outflow of mass and energy to and from their environment, are never in their lifetime in a state of thermodynamic equilibrium, yet are able to maintain a constant state. Many of the characteristics of living/open systems, which seem paradoxical in the light of the laws of physics could be accounted for in this manner.¹²

¹² Von Bertalanffy (1973:38-40) elaborates on his point with reference to recent scientific efforts to account for open systems. He briefly highlights two consequences:

First, the conclusion that the final state of a closed system is unequivocally accounted for in its initial conditions (what he terms the principle of equifinality). He gives the examples of a planetary system in which the positions of the planets at time t are unequivocally determined by their position at time t_0 , and that of the final concentrations in a chemical system that are dependent on the initial concentrations of the reactants. If the initial conditions or the process were to be altered, the final state would be altered accordingly.

Secondly, the apparent contrast between animate and inanimate nature was highlighted. Hence the law of dissipation in physics, which seemed to be diametrically opposed to the law of evolution in biology. According to the second law of thermodynamics the general trend of physical nature is toward a state of maximum disorder and the leveling down of differences, which would end in the "heat-death" of the universe, where all energy will be degraded into an evenly distributed heat of a low temperature and the process comes to a stop. By contrast, the living world is characterised by embryonic development and evolution, and the trend is towards order, heterogeneity and organisation.

Prigogine asserts that without non-equilibrium and the processes that result from it, the universe would have a completely different structure and that there would be no appreciable amount of matter (1984:231). Thermodynamics is, for him, the first attempt of physics to address the complexity of nature, where the passing of time brings about ever greater complexity in the form of structures, but also degradation and death (1984:129).

As soon as time, interaction between the components of a system and the environment in which a system operates in are brought into consideration the linear picture of the universe changes somewhat. We have seen that the second law of thermodynamics introduced “the arrow of time” back into physics by describing irreversible processes that are dependent on the direction of time (Prigogine1984:12). As time passes, entropy increases and reversibility becomes improbable. The superposition principle breaks down in non-linear systems and it doesn't make sense to talk about the individual behaviour of the elements of the system – the properties of a non-linear system differ qualitatively from the properties of its components. For the system to exist the elements necessarily interact amongst themselves and with the environment.

Non-equilibrium structures need a constant source of energy to maintain them. Because these systems are sustained by a continuous dissipation of matter and energy, Prigogine calls them “dissipative structures” (Kauffman1995:20). Kauffman characterises all living systems as dissipative structures. In complex systems symmetry is broken and the resultant asymmetrical structures ensure a rich level of interaction among components of the system and also create competition for resources. Non-linearity is a necessary, but not sufficient, precondition for complexity (Cilliers 1998:120).

5. Emergence

The concept of emergence is central to theories of complexity. As we have seen the law of superposition does not hold in descriptions of these systems. The argument is that there is more to certain systems than is evident

from examining the individual parts that make up the system. This something “more” – these extra characteristics which emerge in the system, but which are not attributes of its individual parts – is what is meant with the concept of emergence. Certain properties become evident when a system is viewed as a whole - they arise as a result of the interactions of the elements that comprise a system. Interaction (of the elements amongst one another, and also of the system with the environment) generates emergent properties.

John Holland states that he is unable to give a concise definition of emergence, citing the complicated nature of the phenomenon as reason (1998:3). His inability to define the phenomenon does not prohibit him from writing extensively on it, and he is adamant that emergent properties should not be regarded as mystical or even rare phenomena. Holland (1997:11) calls emergence a “pervasive phenomenon”, which can be seen in domains as diverse as seeds, scientific models, ethical systems, the evolution of nations, the spread of ideas and board games – the latter being a useful analogy for explaining emergence.

Holland uses chess as an example for explaining the rudiments of the concept: Firstly, even though the game has fewer than two dozen rules, there are a vast number of possible configurations and moves within legitimate play. Furthermore, it is impossible to glean a complete picture of the progress of a given game by simply adding up the values of the pieces left on the board. How the pieces interact and support one another and their positions on the board are just a few of the important factors that need to be taken into consideration when assessing the status of a game (1997:32), and there can be many more. This example is, of course, one of a comparatively simple system, where there is no sign of factors like self-organisation, adaptation, etc. What it does do is provide us with a picture that could amount to a broad definition of emergence as an instance where: “a small number of rules can generate a system of surprising complexity” (1998:3). What is generated is not the complexity of random patterns: the resultant systems exhibit recognisable structures that are dynamic – they change over time. The rules that generate the system stay invariant, but the things that they govern are in flux. In other words emergent phenomena are typically persistent patterns

within a system with changing components. Relatively consistent rules (or laws) generate complexity and the flux of patterns that follow lead to “perpetual novelty” and emergence.

Another simple example to illustrate what is meant by emergence is that of phenomena such as freezing and evaporation, or phase transitions, which are some of the most familiar examples of emergent phenomena exhibited by physical systems. Matter changes from one structure state to another (for example water to ice) in a way that is not discernible from the parts (water molecules). The example of the phase transitions of water molecules is an example used by John Stuart Mill in his treatment of the subject of emergent phenomena. He comes to the conclusion that through studying the structures and properties of oxygen and hydrogen separately would not enable a theorist to deductively infer that, together, they produce water. Mill gives the following criteria for emergence to occur (Auyang 1998:173-174):

1. The emergent character of the whole is not the sum of its parts
2. An emergent character is of a type completely different from the character types of the constituents
3. Emergent characters are not deducible or predictable from the constituents, investigated separately.

Emergence belongs to the structural aspect of the system. The system does not need to have certain kinds of constituents or mechanics to have emergent properties, hence many very different kinds of systems can exhibit emergent properties. Again, these systems are not limited to physical systems and can include social, cultural and biological systems as well. In fact, Auyang warns that physical systems are relatively simple in comparison with social systems, and that the social sciences should consequently take note of the experiences in the physical sciences and be wary of simplistic connections between micro- and macro-descriptions (193). She highlights the difficulties involved in treating emergent phenomena theoretically: in systems that are

thoroughly interconnected, slight perturbations propagate through and affect the entire system, which can result in a system with behaviour that can be “multifarious, unstable and surprising” (1998:183).

In biology the concept of emergence is used by theorists like Kauffman to explain the existence of life as we know it (1995:23).¹³ The main thrust of this (although be it controversial)¹⁴ argument is that the order that we perceive in the living world is an expression of underlying, fundamental laws and rests on the presupposition that a collection of sufficiently complex molecules can aggregate into systems which we would call “alive”. What we would call life would not be a property of a single molecule, but an emergent property of a system of interacting molecules (24).¹⁵ Here “alive” can be understood in the sense of being a system that can sustain itself, evolve and reproduce.

Kauffman is sceptical about the prevailing evolutionary theory that sees life as arising from a primordial soup and then evolving toward ever greater complexity through a succession of hits and misses. He argues that metabolic networks (and thus life) emerged from the primordial soup “full-grown” – already containing the minimal requirements for life. He bases his hypothesis on the premise that the simplest thing known to be alive, *pleuromona*, contain the minimal set of genes for something to be what we would call alive, and that this minimal complexity can be likened to a kind of “natural phase transition in complex chemical systems” (1995: 33-48).

The possibilities of emergence are compounded when elements of the system allow for some capacity for adaptation and learning. Some of the fundamental aspects of this kind of agent-based emergence are aptly captured in the metaphor of an ant colony. The individual agents (ants) have a limited repertoire, but the colony as a whole exhibits remarkable flexibility in its interaction with its environment and exhibits emergent behaviour which

¹³ The same idea runs through “Emergence”, an article in which Holland (1997) asserts that: “we will not understand life and living organisms until we understand emergence” (12).

¹⁴ Refer to Auyang 1998: 202-203.

¹⁵ This view contrasts with that of Henri Bergson, and of a substantial chunk of the Western philosophical tradition, which proposes an *élan vital*, an insubstantial essence animates the inorganic molecules of cells and thus brings them to life.

outstrips that of the individual agents (Holland1998:5). One of the most important characteristics of such agent-based emergent behaviour is that there is no direction by a central executive (something analogous to Dennett's (1991:101-138) *Cartesian Theatre*¹⁶). The idea of ordered behaviour, without a central executive agent, will be explored in greater detail when the concept of *self-organisation* is discussed.

The example of the ant colony also serves to illustrate the limitations of *averaging* when studying the behaviour of large numbers of agents, where traditionally individuals are taken to exhibit typical or "average" behaviour. The overall behaviour of a group of agents is then seen as the sum of the average behaviour, but, as Holland argues, the interactive behaviours of the ants bring a coherence to the colony that cannot be predicted through summation. Consequently we have an illustration of the most important aspect of emergent properties: they are the product of interactions between agents (nodes) and dependent on context. The context in which an emergent pattern arises determines its function (Holland1998: 121-226) - this characteristic will be explored in greater detail later.

6. Framing and modelling

A recurrent theme when studying complexity is that of its dynamic nature and the difficulties that this nature poses to studying and theoretically encompassing complexity. Whereas closed systems can be systems with very few variables and thus lend themselves to formal description, open systems are often too large and too complex. In studying an open system we always have to contend with successive states of the system, with numerous and diverse possible historical processes that have led up to a certain state and with numerous possible states of configuration in the future. Emergent phenomena occur within a greater system with a myriad interacting components and it is impossible to take all relevant components into consideration and to keep track of all the effects of their interactions. At the

¹⁶ Cf. pp. 128-129 for a discussion on Dennett's *Cartesian Theatre*.

same time it must be remembered that the system is dynamic, and that the analyst has to deal with perpetual novelty in the system. If the basic mechanism of the system allows for adaptation and learning, a further dimension of difficulty is added to any attempt to model the entire system.

The fact that open systems do not lend themselves to complete formal description is not due to practical constraints (like the lack of knowledge or resources), open systems are not fully formalisable in principle. The reason for this is that open systems have no fixed boundaries. It is not possible to determine with certainty which components belong solely to the system and which do not. On top of that, we have seen that open systems change continuously and in unpredictable ways. Components are added and detracted and the relationships among various components change, which, in turn, influence the rest of the system. Open systems do not operate according to the dictates of a *telos* – changes are unpredictable, irregular, and contingent. Coupled with the characteristic of decentralised decision making, where no specific component controls the system, and with changes that are the result of many factors, including changes in the environment and in other distant parts of the system, we end up with a mercurial, changing system that cannot be pinned down long enough to be described. With the impossibility of constructing a static model of the entire system and all its possible states and configurations, it becomes necessary to determine the appropriate level of detail at which the system will be scrutinised and the mechanisms of the system which are relevant to that particular study. As mentioned earlier, Cilliers calls this process “framing”¹⁷.

A characteristic of the process of framing, which has important implications for the study and modelling of complex systems, is that a frame is an abstraction. Some aspects and characteristics of the system are abstracted from the rest of the system and given a special status – they are not natural entities and they are not part of the system. We impose frames on entities in our descriptions of them (Cilliers 1989: 40). Not that the argument

¹⁷ Refer to Cilliers (1989: 36-43) for a brief discussion of framing as pertaining to the analytical/empiricist domain and the applicability of this concept to complex phenomena.

here criticises or calls for the abolition of the practice of framing – that is impossible. Without framing communication would be impossible, we need to frame in order to give a description of something.¹⁸ But, in the light of the complex nature of most of the (open) systems that we encounter, complexity theorists would argue that comprehensively framing these systems is, in principle, impossible and that any frames that we do impose on the system will necessarily have to be of a provisional nature. To quote Cilliers (1989: 42):

But we have to frame. It is the imposed frame that creates the levels, the depth in which we can operate, can draw distinctions, can make oppositions, can work...It is the imposed frame that creates the safe space where there can be talk of truth, rationality, causality, purpose and constraints. As long as we remain well inside the clearing, our work can continue, but as soon as we start exploring the limits of our discourse, we are caught in the vortex of the logic of the frame...to have a theory is to draw a frame.¹⁹

To have a theory is to draw a frame – this statement calls forth all kinds of implications that will be expanded upon throughout this study. Most notably that of our active participation in our perception of the world, through constructing models of it, and the necessity of recognising the nature of, as well as the implications and limitations of these models.

Holland (1998:13), for instance, discusses the process of modelling a complex system at some length and places great emphasis on the active participation of the theorist in selecting detail that he/she considers to be relevant, thus highlighting the provisional nature of our models of the world. He insists that in constructing a model of a system we do not have the benefit

¹⁸ Refer to Cilliers (2001: 135-147) for a discussion of the necessity of the presence of constraints, whether physical or informational, in complex systems. Also see (2000b: 40-50) where similar ideas are developed in the context of rules.

¹⁹ Jacques Derrida extensively explores the ideas of the limit of discourse – aporia – and the logic of the frame. Cf. Culler (1994:193-199) for a concise and insightful account of the logic of the frame, as developed by Derrida in various works. Also see Drucilla Cornell (1992) where she discusses these issues as they apply to law and justice.

of preceding study or hindsight to guide us - the relevant laws have to be *selected* in the process of constructing a rule-based model of a complex system and deduction and derivation play a limited part at the time of selecting the rules and constructing a model. Our attention is attracted by a recurring pattern in a particular system (Holland uses the example of the orbits of the planets), and through a process of induction – moving from particular observation to abstract description – we construct a model of the selected phenomena. Only recurring patterns in a given system will lend themselves to being observed and considered to be part of the mechanics of the system. Knowing which details to ignore and which to include when constructing a model is a matter of experience, rather than a matter of deduction. If all goes well the end result will be a description of repeated elements that suggest rules or mechanisms, according to which the system operates. Not being able to reduce the behaviour of a complex system to a set of basic laws does not mean that it cannot be modelled or studied: Holland suggests that one *can* reduce the behaviour of the whole to the *lawful* behaviour of the parts, if one takes the non-linearity of the interactions into account (122).

7. Self-organisation

Part and parcel of the dynamic nature of complex systems is their remarkable tendency to display organisation. We have already touched upon this feature in our discussion of emergence: random patterns resulting from the interaction between components in a system do not make for emergent properties; emergent properties are ordered and *recurring* patterns that come about through some kind of organisation among the components of that particular system. Self-organisation²⁰ is an emergent property of complex systems, and is a result of the interaction between local components. We call

²⁰ In much of the literature on complex systems one comes across the term “autopoiesis.” The phenomenon that the term usually denotes is akin to the phenomenon of self-organisation. And for the purposes of this study and with the intention to avoid needlessly multiplying terms we have elected to omit any reference to autopoiesis and to add its denotation under that of self-organisation.

to mind Holland's example of the ant colony. This agent-based model is useful when trying to make sense of the ability to self-organise. The ants, despite their limited individual capabilities, display remarkably coordinated behaviour, which leads to emergent behaviour in the colony as a whole, with effects far surpassing those that the individual might be thought to accomplish on its own. What is distinctive about organised behaviour in a complex system is that it is not directed by some central, executive agent. The individual components or nodes of the system react to information available to them locally, which translates into complex and organised behaviour on a systemic level. Any one node is not, and cannot be aware of the behaviour and structure of the entire system at any one time, which itself relies on coupled behaviour, contingencies and is subject to continuous change.²¹

There are many advantages to not being directed by a single centre of control, but a system must, by definition, display coordinated and interactive behaviour. The idea in the complexity sciences is to explain how a complex system, like that of the biosphere or language, can attain such a high degree of order, without needing to postulate some external designing or directing agent or some form of executive internal control. It is important to stress that the phenomenon of self-organisation is a profoundly pragmatic - perhaps one could even say prosaic - occurrence, which has to do with the optimal functioning of a system and which does away with the need for recourse to metaphysical explanations when trying to account for the morphologies of phenomena.

Cilliers (1998:12) describes self-organisation as "a process whereby a system can develop a complex structure from fairly unstructured beginnings." In keeping with views raised in the last paragraph Cilliers sees the significance of self-organisation in the idea that internal order can come about

²¹ Birute Regine and Roger Lewin (2000) from their article *Leading at the Edge: How leaders Influence Complex Systems*, in which they apply the insights gained from complexity theory to organisational and management theory, use the term "organizationally flat" where they suggest that companies that are less hierarchical in the managerial approach will be more flexible and adaptable and thus better suited to keep up with the fast-changing face of business. This approach can be misleading in that it contains the implication that complex systems do not contain hierarchies and are in some way flat. See Cilliers (2001:142-145) where he argues that complex systems do in fact (and necessarily) contain hierarchies.

without the need for an external designer or an internal form of centralised control (88). We will later return to what exactly the significance of these characteristics would be. In a complex, adaptive system the relationship between the components of the system is altered both by the system's interaction with its environment and by the system's own history. These provide limiting factors (constraints), which contribute to the order that a system settles into.

Per Bak, one of the theorists that introduced the concept of "self-organised criticality"(1996:1-3), aims to describe the tendency of large systems with many components to evolve to an unbalanced, yet structured state – a "poised" or "critical" – state, where minor disturbances may lead to events of all sizes. To illustrate how most changes to the system take place through catastrophic and unpredictable events, rather than gradual change Bak uses the image of a person trickling sand through his/her fingers to form a pile. As the pile grows, little sand-slides occur, which become bigger as more sand is added to the pile. The pile reaches a critical point where no more sand can be added to the pile without causing sand-slides, and where the size, frequency, and the impact of the slides are inconsistent and unpredictable. The same mechanisms that cause small slides (minor changes) can cause major sand-avalanches. The behaviour of the sand pile can no longer be understood in terms of the individual grains of sand and develops a dynamic that relies on the configuration of the whole system. The grains have collaborated to form an interacting system, which, when reaching a certain critical point, undergoes unpredictable changes of inconsistent magnitude and frequency.²²

We have already mentioned the evidently pragmatic nature of self-organisation, in that the theory is that self-organisation ensures the optimal functioning of the system. Cilliers (1998) explains what takes place in self-organisation as the system organising itself to a critical point where single

²² This is of course a simple example that does not do away with an external agent (the person trickling the sand). Its significance lies in its ability to illustrate a point of saturation for a system, after which the system undergoes changes that are the result of a combination of its own configuration and environmental factors.

events will have the widest possible range of effects and where the system can obtain optimum sensitivity to external inputs. These critical points are also called “attractors” – states that a system eventually settles into, and which are determined by the properties of the system (Lewin 1995: 17).

Cilliers (1998:97) provides us with an example with which to visualise the behaviour of a system when it settles into an attractor state. Firstly we need to visualise the state-space of the system. The state-space has a separate dimension for every independent variable of the system. Thus a system with three variables – temperature, volume and pressure – has a three dimensional state-space. A system with a network of 1000 variables (or nodes) has a state-space that is a thousand dimensional (97).²³

Every possible state of the system is characterised by a unique point in the state-space. The dynamics of the system (the way in which the system unfolds through time) forms trajectories through the state-space. When a number of trajectories converge on a certain point in the state-space that point is called an *attractor* - a stable state of the system. A stable system has only a few strong attractors, whereas an unstable system would have no strong attractors (but can have many weaker ones) and would jump around chaotically. A chaotic system has no structure and is useless, while a stable system with too *few* attractors is very rigid and cannot readily adapt to changing conditions. To return to the idea of self-organised criticality, it postulates that a system will balance itself at a critical point between rigid order and chaos. At this critical point single events in the system can have the widest possible range of effects, without disrupting the system – the system will be at its most sensitive to external input, without being unstable. The system will also be able to change its state with the least amount of effort. The

²³ Auyang defines state-space as a “structured collection of all possible momentary states of the individual” and argues that something changes substantively when its states have different characteristics at different times (1998:215). Successive states taken as a whole constitute an entity’s history or the process that it undergoes to reach a certain state at a certain time. In accordance with the idea of coarse-graining and framing, Auyang argues that the impossibility to encompass an entire system (including its history) in state space, we must resort to narrative explanations when describing the system.

tendency that a system has to move to criticality leads to an increase of complexity in the system (97-98).

As already mentioned, theorists contend that economy and flexibility underlie this behaviour of complex systems (Kauffman 1995:26; Cilliers 1998:98). The ability to self-organise is an intrinsic characteristic of a complex system, which ensures that the system is organised in such a way that it functions at optimum capacity and to ensure that resources available to the system are fully utilised. The absence of rigid hierarchy and central control has the important effect of allowing the system to be extremely flexible and adaptable.

A simple and useful example is that of Bénard cells as described in Juarrero (1999:119-120). We start off with a pan of shallow water at room temperature, where the identical water molecules randomly bump into one another. As we uniformly heat the pan from below, the liquid initially stays at a uniform temperature and the system is in thermodynamic equilibrium. Eventually, however, the bottom of the pan grows hotter and a temperature gradient occurs in the liquid, which increases thermal non-equilibrium within the system; the system becomes unstable and is unable to maintain its current "organisation". The independent water molecules reorganise themselves into a system that is macroscopically orderly; patterned into "rolling columns of hexagonal cells, called Bénard cells"(120). The dynamics of the system switches from heat conduction to heat convection; the system has undergone an abrupt transition from chaos to order, in order to facilitate the critical temperature gradient (non-equilibrium state) that had built up. Now context (the environment) plays a significant role in that the behaviour of individual molecules becomes dependent on that of surrounding molecules and on the temperature to which the water is heated. The control parameter (increased heat) drives the system away from equilibrium to a phase transition (bifurcation). In this example the heat is controlled from the outside and if it is reduced the Bénard cells will disappear, if it is increased the water will become more turbulent (turbulence being highly complex, ordered behaviour). Systems that can adjust their own control parameters and thus reorganise themselves are truly self-organising systems.

It is the concept of self-organisation that enables Stuart Kauffman to question existing theories on evolution. In a series of idealised experiments with virtual networks Kauffman has demonstrated the seemingly uncanny tendency for systems with many different interacting components to move to certain attractors in the system and settling down into a recurring and recognisable pattern of behaviour. As can be imagined, this kind of experiment can hold some exciting implications for the biologist in as far as the evolution of biological systems are concerned.

As a metaphor to explain how “autocatalytic metabolisms”- self-organisation among molecular species in the primal waters - came about, Kauffman uses the image of a network of light bulbs, wired together at random with an electrical circuit – this serves as an apt example for all manner of complex systems (1995:71-79). In this case, one molecule catalysing another could be thought of as one light bulb switching on another one. The aim is to demonstrate that, under the right conditions, this random network will not, as might be expected behave in a random manner, but that, with time, order will arise spontaneously from the interaction between the bulbs. The bulbs settling into a coherent pattern of switching on and off will be analogous to trajectories in a system converging on certain attractors in its state-space.

Each bulb can have only two values, on and off, depending on input from other bulbs. Nevertheless, the number of possible states in such a network is vast; all the bulbs might be on or off, or any number of configurations combining these two states in between. As we explained earlier, this range of possible configurations is called the state-space of the network. Assigning *Boolean* functions²⁴ to each light bulb - meaning that each bulb will, in its next state re-act to the input from the current state of the other bulbs that it is connected to - causes the bulbs to behave in an interactive manner. As the network proceeds through a series of states each bulb (node; gene; molecule) examines input from the other bulbs and then switches on or off (becomes active or inactive) according to the rules assigned to it for

²⁴ Named after the nineteenth century mathematical logician, Charles Boole.

reacting to other signals. The network then proceeds to the next state and the process is repeated. A system, consisting of a finite number of bulbs will have a finite number of possible states. Once started off, the system will run through a sequence of states, or a trajectory. Since, with a limited amount of bulbs, there are a finite number of possible states, the system will eventually hit upon a state that it has previously encountered and will run through the associated trajectory again. The system is deterministic, in the sense that the behaviour of light bulbs is constrained by the behaviour of those that they are connected to, and will eventually run around a recurrent loop of states - a state cycle - where it will repeat its pattern of on/off bulbs again and again. A smaller network will have a smaller number of possible states, which would make it easier for an observer to detect a repeating pattern denoting a state-cycle. A large network, with more elements and a vastly bigger amount of possible states may have a number of possible states in its state-cycle so vast that it could conceivably take so much time to run through its cycle that we might be unable to detect a pattern, as the system will never repeat itself in our lifetime. The behaviour of such a network will appear to be utterly random to us. Networks that settle into small state-cycles exhibit, what seems to us to be, repeatable behaviour.

We have already seen that attractors are a source of order in dynamical systems in that different trajectories converge on this point in the state-space where they are “trapped” in this sub-region of the state-space, and that both Cilliers and Kauffman cite small attractors as a prerequisite for order. Complex systems are often characterised by what has been named *strange-attractors*. Strange attractors describe ordered patterns, which will still allow individual behaviour to fluctuate. So even if trajectories in a system are caught in an attractor basin, their behaviour is not so rigidly constrained so as to cause the individual trajectories to repeat a phase exactly. Consequently, even though the system is constrained in a state-cycle and does display order, it is still hard to discern an overarching, repetitive pattern of order, as individual trajectories are never exactly identical, but approximate. As Juarrero has it: “The width and convoluted shape of strange attractors imply that the overall pathway they describe is multiply realisable” (Juarrero 1999:

155). Even though the trajectories in these systems appear to be random, such intricate behaviour patterns are in fact an indication dynamic, complex and context-dependent organisation (154-156).

Another prerequisite for order in a dynamical system is that of homeostasis (Kauffman 1995:79): a system needs to be resistant to small perturbations. If we arbitrarily choose a light bulb and switch it off and on again, the system needs to be able, more often than not, to return to the state-cycle from which it has been perturbed, in order to be stable. Not all systems possess homeostatic stability – a system with too many attractors is unstable, and any slight perturbation might send it into another basin of attraction, thus changing its cycle and disrupting the system's "pattern". A system in which all its attractors are unstable in this way would be a chaotic system, vulnerable to all manner of fluctuations and never able to repeat its cycles through the state-space, thus never able to retain its order.

Kauffman uses the insights gleaned from his experiments on Boolean networks and especially his computerised experiments²⁵ to amend the theory of evolution. As we have mentioned earlier in this chapter, Von Bertalanffy criticises the Darwinian theory of evolution, where organisms appear to be the random products of chance and haphazard mutations. It is this aspect of the theory of evolution that Kauffman seeks to address. He is of the conviction that different cell-types are equivalent to state-cycles in Boolean networks, the later being an accurate, if idealised, model for the generation of cell types. In his own words: "I suspect that the fate of all complex adapting systems in the biosphere – from single cells to economies – is to evolve to a natural state between order and chaos; a grand compromise between structure and surprise" (1995:15). His aim is then to show that order arises naturally and that much of the order in organisms is not the result of natural selection, but the result of spontaneous self-organisation within systems.

Systems exhibit their most interesting behaviour when they operate at, what is often called "the edge of chaos"- poised on the edge between order

²⁵ Refer to Kauffman (1995), Lewin (1993: 48-139), Ayuang (1998) for extensive discussions on these experiments.

and disorder, akin to the point between phase transitions in physical systems (26). At this point networks are best adapted to handle complex calculations, yet are still controllable. Networks also maximise their capacity to evolve, responding to minor changes without becoming chaotic. Most minor changes in the system will have a minor effect on the system, with it still being able to return to its attractor state. Some mutations may, however, have a greater effect on the system, and will lead to evolutionary changes. Auyang elaborates: "...networks at the edge of chaos are able to evolve both by accumulation of small changes and by dramatic changes that in which evolutionary novelties emerge. The emergent changes might be triggered by a random mutation or by a change in the environment. The conditions under which natural selection is most powerful are also those in which self-organisation and historical contingency are most likely" (1998:202). Kauffman claims that these theories of self-organisation and selection are the laws and general regularities that underlie evolution generally. His views are not uncontroversial and his idealisations and computerised models are accused of being too far removed from empirical evidence and are found to be quite alien to many mainstream biologists (see, for example, Auyang 1998: 202). Yet, his ideas are compelling and may prove to have many interesting philosophical implications.

8. Feedback, the environment, and adaptability

The bulk of the discussion on self-organisation so far has been to indicate how complex systems acquire structure. We have said that one of the most important characteristics of this structure is that it comes about as a result of the dynamics and interaction between the components in a complex system, and the system and its environment, no external designer or pre-programmed "software" is required. The preceding section links to the first half of Cilliers' definition of structure, while touching on a second characteristic which needs to be taken into consideration for us to have a complete picture of complex structure: "The notion of 'structure' pertains to the internal mechanisms *developed by the system* to receive, encode, transform and store information on the one hand, and to react to such information by some

form of output on the other" (1998:89; my emphasis). Cilliers, more so than the other theorists discussed thus far, places emphasis on the role of the environment in the development of structure within a system. The environment presents a number of constraints, which curtail and influence the development of structure. The resulting structure is the product of a complex interaction between the environment in which a system finds itself, the present state of the system and the history of the system.²⁶ In accordance with Kauffman, Cilliers explicitly defines the capacity for self-organisation as a property that enables a system to change its structure adaptively, which allows it to "cope with, or manipulate its environment" (90). Here we touch upon a further aspect of self-organisation: a self-organising system not only reacts to and adapts itself to its environment, it also interacts with the environment and has the ability to instigate changes in its environment.

There exist causal relations between the system, its environment and the system's history, referred to as *feedback*, which explains how the system is shaped by its environment, but also how the systems shapes its environment. Open systems that exchange matter and energy with their environment, receive feedback from that environment. As Juarrero would have it: "feedback embeds them in that environment in such a way that they are simultaneously context-dependent and initiators of behaviour" (1999:75). In a non-equilibrium system any naturally occurring random fluctuation can become amplified, with the same kind of effect as, in our earlier example, the increased heat had on the Bénard cells.²⁷ The system moves to dissipate the

²⁶ "History" here refers to already actualised states of the system. A certain configuration of the system – which is the result of previous states of the system and environmental constraints – already precludes some possible states, while enhancing the probability of the actualisation of others. History in this sense should not be understood as a chronological series of major events. As Cilliers put it: "The history of the system is contained in all the individual little interactions that take place all the time, distributed over the whole system (2000a: 25).

²⁷ A simple example by which to understand feedback is that used by Peak and Frome (1994:5-6), namely: audio feedback:

Picture a speaker and a microphone connected to a stereo. If the microphone is pointed towards the speaker and the volume on the stereo gradually increased, the speaker will soon start making an electronic squealing noise, as sometimes happens at a public speaking or at concerts. The reason this happens can be explained as the result of an iterative process: feedback. The speaker is designed to convert an electrical signal into vibrations in the air - sound. The microphone is designed to convert vibrations in the air into

non-equilibrium that builds up and hence in each re-organisation becomes better able to deal with the environment in which it finds itself. The system can be driven into a new dynamical organisation. In this manner, self-organisation can, by means of its own dynamics, cause the system to evolve (Juarrero 1999:122). The fluctuations in the system are random and contingent and because they serve as the nucleus around which the re-organisation occurs, "...the progression of those evolutionary processes also marks the trajectory of an increasingly individuated system"(122). These self-organised systems derive their identity from the organisation of the processes that constitute them and not from the primary material of their components (124). The system becomes context-dependent in that its constituent elements become dependent on the behaviour of their neighbouring elements as well as what happened previously in the system. Feedback loops incorporate time into the system, by making the system dependent on its history - by incorporating the system's past into its present structure, feedback adds the additional constraint of already actualised states and past experiences influence the system's subsequent behaviour (138).

Open systems do not passively serve as conduits for energy. Although the open system imports energy from its environment, the energy, by virtue of the systems dynamics, is diverted to maintain the system's internal organisation. Self-organising systems are self-referential, and new components are "accepted" into the system by virtue of their ability to enhance the overall organisation of the system. The system's organisation makes for what Juarrero calls "an internal selection process", which is established by the system itself, and operates to preserve and enhance the

electrical currents. The microphone picks up background noise in the room and sends an electrical signal to the amplifier, which amplifies the sound that is then emitted by the speaker. This adds to the noise in the room that is subsequently picked up by the microphone. The stronger sound is converted into a stronger signal to the amplifier and the speaker in turn; inevitably the speaker emits an even louder noise. After additional passes through the loop, the microphone squeals violently: this repeated loop in which the system causes itself to increase the volume of its output is an example of feedback. (The same experiment can be conducted with a camcorder and a video monitor, with intricate visual results.) This is of course an example of a simple, linear system. Feedback in complex systems will inevitably be a more complex affair.

system (126). This ability is what we have in mind when we say that self-organising processes are primarily informational. The components that are imported into the system are selected through the internal dynamics of the system, based on the system's internal requirements. Under pressure from its environment, the system seeks to enhance its ability to process energy and matter flows, and thus enhancing its cohesion and integration and functioning capabilities as a whole. Again, the system does not organise itself in terms of some kind of *telos*, in the Aristotelian sense of a final cause, but in terms of the interaction between dynamics of system and contingencies of its environment. These contingencies make it impossible for the system to evolve along some previously established path. In fact, it is the very contingencies that systems have to contend with that make the capacity to self-organise an essential tool of survival. Contextual constraints do not only hold for biological and physical systems, but for informational systems as well. It is perhaps prudent to explicitly state that contextual constraints are a necessary precondition for the establishment and maintenance of complex systems. This is an important concept to grasp and merits brief explication.

9. Information and enabling constraints

Complex systems must have the capacity to encode and remember information about their environment, which enables them to cope with changes in the environment. We have seen that time is brought back into the equation when dealing with complex systems, in the sense that a system's history influences its present structure as much as its environment. In contrast with thermodynamics – which we have seen discredits hypothesis that retrodiction of the trajectory of a system is possible, but which still allows for prediction of future states of the system (thermodynamic equilibrium) – complexity sciences, in dealing with open systems, discredit both retrodiction and prediction of the trajectories of such systems. The reason that both are discredited lies in the fact that there are just too many relevant variables and too many possible states of the system that are never actualised. The possibilities that are actualised, the attractors that the system settles into, are established partly by the environment (and the function of the system in that

environment), and partly by the history of the system, by states that have already been actualised.²⁸ This interaction of the system with its environment need not be physical; it can also be thought of as the transference of information (Cilliers 1998:3).

Information, as it pertains to complex systems in general should not be understood as *meaning* in an overtly anthropomorphic sense. Information has its role to play in all instances of self-organisation in complex systems, and what is meant by information depends to some extent on the system under discussion. When we come to our discussion on information as it features in neural networks and language, the link between information and meaning becomes vitally important.

When we say that the system is constrained by its context, it is important not to create the impression that these constraints are negative, in the sense that their role is only that of inhibition and limitation. Constraints in complex systems also play an enabling role; the term “system” implies structure and constraint. Constraints also enable the system to maintain its identity. Without these constraints there would be no system.

A useful analogy by means of which to explain the function of physical constraints – which arise out of self-organisation in complex systems - is that of information as depicted in communications theory (Juarrero 2000: 34). In this contexts constraints are what make communication possible. Even though, in utterly random signals where possible meanings are equiprobable, the potential of possible meanings is at a maximum, these signals can have no informational content. In a situation where possible meanings are completely random (chaotic) one could say anything, but would in truth be saying nothing – the unconstrained disorder makes it impossible to detect patterns, which would equate with a message and one ends up with

²⁸ Here again we call to mind the example of the Boolean networks of light bulbs where each successive state of each bulb (node; element) – whether it switches on or off – is determined by its previous state as well as the previous states of those bulbs that it is connected to. The previous states of the light bulbs can be seen as analogous to the history of the system, and the Boolean functions ascribed to each bulb can be seen as analogous to the system's self-organised structure that arises as the result of its function in the environment.

undifferentiated white noise. In the words of Juarrero: "When anything is as possible as anything else, and nothing is connected to anything else...nothing can signify or communicate anything" (Juarrero 2000: 35). By contrasting significant signals with "background noise" and subjecting the way that the signals can be arranged and encrypted to certain rules, one reduces the randomness and allows patterns to emerge. The same principle applies in language: contextual constraints that ensure the interdependence of words and sentences - grammatical and syntactical rules - allow a structure that lends itself to meaningful interpretation. If letters were random, if any letter could show up at any time, how would it be possible to communicate anything? Standardised language rules cause the number of ways in which the components of language (letters, words, sentences, etc.) can be arranged to be drastically reduced – instead of inhibiting the informational capacity of language, however, these contextual constraints drastically increase it. The synchronisation of the individual components (letters in this example) causes number of possible arrangements to be reduced, but the possibility of transferring encodable information to be greatly increased.

This (very brief) description of the mechanisms of language (a complex system), as far as its capacity for informational content is concerned, is analogous to the working of constraints within physical complex systems. We have seen that, far from being in equilibrium, complex systems are interdependent and context-dependent through being constrained by the behaviour of neighbouring elements (molecules/agents/nodes), previously realised states, and the influence of the environment. The dependence on other elements etc. makes for a system, the behaviour of which is ordered or determined by interaction of various factors. As is the case in our example, where constraints reduce randomness and make communication possible, these context-sensitive constraints, far from limiting complexity, make it possible. Juarrero uses the phrase "enabling constraints" (2000:37) and describes the emergence of self-organised systems as the "sudden closure of context-sensitive constraints." The emergence of Bénard cells illustrates the abrupt appearance of an open, interdependent system, far from equilibrium, where the new relationship among the molecules establishes new boundary

conditions for those molecules, which makes for cells with very different properties than those of individual water molecules. An important effect of context-sensitive constraints is that they restrict, or should we say regulate, the flow of energy and matter between the system and the environment. The organisation of the self-organised system determines the stimuli to which it will respond (39). Because the elements are interdependent, the behavioural variability they would have had as independent elements is constrained - this enables the system to preserve its organisation and its identity as a whole. Instead of some governing component determining and regulating the behaviour of the system, the relational whole of the system, governs the behaviour of the system.

10. Conclusion

This brief overview of complexity theory is inevitably simplified and rudimentary. But hopefully it will prove to be adequate for our purpose in this work, which is to model the self on the principles of complexity theory. "The self" is a polymorphous concept that has undergone many metamorphoses in the Western philosophical tradition, which means that such an enterprise is anything but straightforward. In the interest of both orienting ourselves with regard to the self *and* (relative) brevity the next chapter will trace the development of the concept roughly as it parallels developments in the history of science. Hopefully this will provide us with a conception of the self that will be delineated and viable enough to serve as a basic understanding of what is meant by "the self" in this work. In subsequent chapters the possibility of applying the principles of complexity theory to the self, and possible consequences of such an endeavour, will be explored.

Chapter Two

Landmarks in the History of Self:

The Legacy of the Enlightenment

It has become to seem that matter, like the Cheshire Cat, is becoming gradually diaphanous until nothing of it is left but the grin, caused, presumably, by the amusement at those who think it is still there.

Bertrand Russell quoted in Popper (1977:151).

Minds are not bits of clockwork, they are just bits of not-clockwork.

Gilbert Ryle (1960:20)

1. Soul, self, consciousness, ego, subject, mind, identity, self-concept, personality – semantic nitpicking or rigorous distinction?

Theories of the self, in one form or another, is one of the most pervasive topics in Western philosophy²⁹, and yet the self seems to have proved an elusive phenomenon. Not only does the entity which we mean to encompass under the concept “self” seem mercurial, but in the vast literature that covers this topic many different terms are used interchangeably and to denote what would seem to be roughly the same thing. What should we understand the terms self, consciousness, soul, ego, subject etc. to mean?

²⁹ Due to my training and lack of knowledge with regard to other philosophical traditions this work is restricted to the Western philosophical tradition and the reader should read any mention of philosophy as referring to the Western tradition.

Do these terms denote roughly the same thing or do they serve to highlight the distinguishing characteristics of very different and very particular entities? Does it make sense to speak of the ontology of the self, or are the logical positivists right in viewing the concept of self as being a linguistic confusion? Although there does not seem to be any rigour in the general application of these terms, the aim, in this work at least, is to delineate clearly what is meant here with the concept “self” (this being the concept that will be employed throughout this work).

One reason for the hodgepodge of terms used to delineate the self is the metamorphoses conceptions of the self have undergone throughout the history of theorising on human being. In some cases the respective terms also seem to be devisable in terms of their denoting a capacity for agency as opposed to a passivity of sorts. Whether one construes the entity in question to be an active and autonomous agent or as a passive epiphenomenon of various processes would seem to partly determine the term one would want to employ.³⁰ What most of these concepts have in common is their reference to that which “animates” the apparently inanimate matter of which our bodies are comprised – in an attempt to explain that which makes us human; that which allows the capacity for reflection and especially self-reflection, and (apparently) willed action.

A sparse definition which could serve us at the outset of our discussion could be something along the lines of describing the self as the “me as *subject*”, as opposed to the totality of objects that I am aware of and that are “not-me”. Our point of departure could then be that we are examining that “something” that enables us to be aware of both ourselves and of the world, and to actively orient ourselves with regard to that awareness. And which, for all intents and purposes, remains a mystery.

This chapter is an attempt to sketch a brief outline of major developments in theories on the self from antiquity to the present day, with emphasis on those that contributed to prominent factors in some current

³⁰ See Armstrong, (1998: 490-491) for a handy, and brief, overview of different mentalist, physicalist and dualist theories on the mind-body problem.

theories on self. As can be imagined, this is a momentous task and far exceeds the scope of this chapter. Much detail and relevant theorists and trends will be ignored, due to expediency, but also due, in part, to the aims of and the rationale for this discussion. The discussion is primarily aimed at orienting ourselves with regard to the origins of many of our conceptions of self, and also to highlight just how much of those different conceptions are based on preconceptions and assumptions that should lose some of their influence as our knowledge of the physiology of the brain – and by implication, the mind – expands and develops.³¹

2. Introducing the self

There by no means exists consensus on the trajectory that theories on the self have followed in the history of the Western philosophical tradition.³² While some theorists regard the self as a ubiquitous philosophical phenomenon, others believe it to be the result of overzealous seventeenth century idealism.³³ For example, both Popper and Rorty attribute the view that the distinction between body and mind is a “new-fangled legend” to Ryle in particular (Popper 1977:151; Rorty 1980:17).³⁴ Popper believes the implication of such a position to be that *pre*-Cartesian philosophy was, on the whole,

³¹ As we shall see, the ease with which this sentence presupposed the interconnectivity of the brain and mind (in terms of the traditional distinction) is not at all obvious to many philosophical positions.

³² It is interesting to note that dualism is not at all a universal, “naturally intuitive” theoretical stance. As Hans Ågren (1998:146) points out, in traditional Chinese philosophy and science (as opposed to Chinese Buddhism) the psychological “was almost never regarded separately from the physiological.” Confucians, for example, disregarded supernatural forces - gods were of no interest, because there was no way of gaining knowledge about them. A dichotomy between body and soul was disregarded for is similar reasons. At the same time the Chinese never developed a psychology of the unconscious.

³³ It is often asserted that the mind-body problem is a distinctly modern phenomenon, a dualism that gained credibility subsequent to the Enlightenment and the works of Descartes and his contemporaries. Conventional wisdom in some quarters has it that the self became prominent in philosophical theorising in the time of the Enlightenment, and that, up until that time it had not really featured as a topic of philosophical discussion.

³⁴ In reading Ryle (1960) it would seem that his position is not as much that the very *concept* of self is a post-Enlightenment phenomenon, but that the identification of the “mental” (and hence the self) as belonging to a logical category other than the physical is a legacy of the Cartesian myth and the “the three centuries of the epoch of natural science” (8). We will discuss this position in greater detail in a following section.

materialistic, which leads him to imply that someone who does not accept a materialistic account of the self/mind has been brainwashed by post-Cartesian, dualist philosophy (1977:151). Popper believes the conception of the self as *soul* to be one of the oldest and most pervasive conceptions of self, and to date back to antiquity (3). Here, the soul is that which is eternal, and which transcends spatio-temporal constraints, something that is inborn, cannot be obtained, and that may, but need not, have the capacity to survive death³⁵.

As with Popper, Levin (1992:3-5) places the origin of the Western conception of self with the ancient Greeks. He believes that both the conception of the inborn soul and conception of the attained soul to have been present from the start of our theoretical tradition. Levin sees Socrates' introspection as his (Socrates') concept of self and likens it to the kinds of theories that would postulate the self as something to be developed and achieved. For Levin Socrates' self is relational, in that it develops through dialogue with others, whereas Plato's conception of self is that of an inborn soul – self-conflicting, divided into reason, drive and appetites. Plato's soul is trapped within the prison of the body.³⁶

Cast in these broad terms it would seem difficult not to believe, with Popper and Levin on an intuitive level at least, that some sort of conception of self is as old as our theoretical tradition. But, the question does arise whether or not the soul-body distinction as conceived in the pre-modern era can be equated with the *mind*-body distinction in the form it has taken subsequent to the Enlightenment. Ryle, for instance, presents a materialist (read: non-dualist) conception of the self as the natural or intuitive position that one

³⁵ Levin (1992) contrasts this ancient Western view of the self as soul with that of various Eastern conceptions, where self/soul has to be attained or achieved (3).

³⁶In *The Republic*, for instance, Plato declares:

But if we want to see it [the soul] as it really is, we should look at it, not as we do now, when it is deformed by its association with the body and other evils, but in the pure state which reason reveals to us. We shall then find that it is a thing of far greater beauty, and shall be able to distinguish far more clearly justice and injustice and all the other qualities we have talked about (1981:444).

See Armstrong (1972:39-43) for a concise discussion of Plato's tripartite soul.

would adopt if one were not influenced by modernist ideology³⁷. While on the other hand, theorists like Popper believe some kind of dualism to be the norm throughout history and, perhaps even a natural or intuitive view³⁸. The question arises, of course, if Popper's "dualism" is not compatible with Ryle's "materialism".

3. An ancient self?

We shall briefly discuss the more remote origins of current notions of soul/self/mind before moving on to the Enlightenment, where the question of self as mind becomes explicit. Subsequently, the Modernist view of the self – as pioneered by Descartes - would become an influential and contentious issue, the reverberations of which can be felt up until the present day debates. Popper does not discredit enquiry into "mind" in some form or another in any way but does believe all philosophers that have a definite position on this issue (pre- and post-Cartesian alike) to be "dualist interactionists". He defines dualism, very broadly, as the tendency throughout history to speak of mind, and soul, and spirit as opposed to the material body. As soon as thinkers like Homer, Democritus and Socrates began speaking of the moral world, mind began to take on a special character, which distinguished it from matter (Popper 1977:153). These Dualists are faced with the difficulty of explaining how the two distinct substances (mind and matter) interact – from there, the term dualist interactionists.³⁹

³⁷ Although Ryle does call the dualist doctrine a myth (1960: 11-24), one is hesitant to conclude that his position is an overtly materialistic one. Ryle's (1960:8) main criticism of Descartes' position is that he perpetrated a category mistake in his conception of mind. In Ryle's own words: "Descartes left as one of his main philosophical legacies a myth which continues to distort the continental geography of the subject... A myth is, of course, not a fairy story. It is the presentation of facts belonging to one category in the idioms appropriate to another. To explode a myth is accordingly not to deny the facts, but to re-allocate them." More about Ryle's criticism and about whether he succeeds in "re-allocating" the "facts," as he deems possible, in a later section.

³⁸ Popper relies on numerous Greek texts and other ancient writings to substantiate his position (1977:149-170).

³⁹ As we shall see Popper (1977: 153) believes that it is the intrinsic difficulties in Descartes' elaborate dualist system that finally lead to alternatives to interactionism.

Popper (1977: 156 -157) speculates that both the development of language and the comprehension of mortality and certain death lead to the conjecture of a ghostlike soul with properties different from that of the body – including the ability to survive death in some or other form. He makes it very clear that he does not imagine this kind of dualism to be a Cartesian dualism, but proceeds to quote various passages from Homer as examples of pre-historic and early historic instances of the mind-body problem.⁴⁰ He goes so far as to cast the *conscious* (read: self-conscious) self as a universal experience of mankind (ibid.). Hence he rejects any attempt to characterise Greek philosophy as being aware of the soul-body problem, as opposed to the “more contemporary” mind-body problem, as a “verbal quibble” (1977:159). He declares the Greek soul to be an entity or substance, which sums up the experience of the conscious self, and thus fulfils a role similar to that of the Cartesian and post-Cartesian mind.⁴¹

Popper believes the source of this verbal quibble to lie in two different views of scientific explanation in the Western theoretical tradition. In his discussion of the history of the ancient self, and its influence on modern conceptions of the self (159-171), Popper emphasises these two opposing views, inherited from the Platonic and Aristotelian traditions. He names these two approaches the “conjectural explanation ” and the “ultimate explanation” and emphasises that both Plato and Aristotle discussed and applied each of the two methods. Conjectural explanation essentially consists of making an assumption (perhaps based on intuition) and then testing the assumption by exploring its consequences. Users of this method are perfectly aware that they can only establish such an assumption provisionally. The method of ultimate explanation, on the other hand, consists of the intuitive grasp of the essence of something or another. Important for the purposes of our discussion is that “intuition” here implies infallible insight, this being a method

⁴⁰ Popper’s argument is that tales of metamorphosis from classical antiquity, of which there are legion, in which the body undergoes fantastic metamorphoses while the self-conscious mind and self-identity stay intact, to be indicative of a conception of mind or consciousness separate from the body.

that guarantees Truth. What one grasps intuitively is the essence and a definition of the essence allows one to explain the phenomenon deductively⁴² (Popper 1977:172).

What is missing, Popper argues, is a full awareness amongst those caught up in this verbal quibble that these two methods differ fundamentally and that conjectural explanation is valid, while ultimate explanation is “will-o’-the-wisp” (173). Popper insists that only ultimate explanations make definite knowledge claims,⁴³ seeing that the conjectural method is self-consciously, well, conjectural.⁴⁴ To Popper it seems that Plato, and many subsequent philosophers, “even Newton”, regard conjectural explanation as tentative and provisional, a stepping-stone to something better (174). Popper goes on to argue that there are two corresponding methods of criticising claims made by these two methods: scientific criticism, which criticises an assertion by examining its logical consequences; and what Popper calls “philosophical criticism”, which criticises an assertion by showing that it is not demonstrable “cannot be derived from intuitively certain premises” (173).

While Descartes and subsequent philosophers argued for (or against) an essentialist explanation of mind, Popper argues that it does not make sense to expect an essentialist answer to our enquiries as to the nature of the mind. He believes that most of the difficulties encountered in theorising on the mind is the expectation that we will discover what the mind is, *in essence*. He points out that we do not know what matter is, in essence, even though we do know quite a bit about its structure and concludes that it makes sense to concentrate on broadening our knowledge on the structure of the mind,

⁴¹ Refer to Popper (1977: 159-171) for a comprehensive account of the history of the soul in antiquity, moving from material soul of Democritus and Epicurus, to the dematerialised mind of Plato and Aristotle.

⁴² Toulmin makes a similar distinction when he explores the influence that the revived Platonic and Aristotelian theories had on modernity. He does not, however, insist on both kinds of explanation to be present in the projects of both theorists. Throughout his book, *Cosmopolis* (1990), Toulmin represents the Platonic approach to knowledge as essentialist and the Aristotelian approach as conjectural.

⁴³ As we shall see is the case with Descartes.

⁴⁴ It is important to emphasise that Popper does insist that there are objective reasons why some theories are to be taken as objectively more preferable than others.

without insisting on taking the supposed essence of mind as our point of departure (174). In short, Popper believes “contemporary” dualism to be the norm when it comes to current conceptions of the relation between body and mind, thanks to the theoretical turn taken in the Enlightenment. His criticism of this essentialist turn is that it lost sight of the fact that the foundational principles on which these theories were built, were, or at least should have been considered to be, *conjectural*.

Richard Rorty (1990:17), although in agreement with Popper in his criticism of some forms of dualism, is, on the other hand, suspicious of the claims that assume “everyone has always known to divide the world into the mental and the physical – that this distinction is common-sensical and intuitive...”⁴⁵ In his words he calls the division between the mental and physical a division into “two sorts of ‘stuff,’ material and immaterial” (ibid).⁴⁶ Rorty entertains the idea that this division, instead of being intuitive and therefore relatively ubiquitous, is in actual fact the propensity and ability to command a particular technical vocabulary, one that had its very origin in the Enlightenment, and in the work of Descartes in particular (22). This assumption leads Rorty to question just why it is that modern theorists associate the phenomenal, or mind “stuff” with the immaterial and matter as material. He finds his answer in Ryle’s observation that we think in ocular metaphors, and that we think of the phenomenal as a “funny kind of particular” before the mind’s eye (31).

With regard to contemporary dualism, Rorty asks why the latter day neo-dualists, as inheritors of modernist dualism, are so sure that the different vocabularies that are used (in philosophy of mind especially) to “describe feelings” and to “describe neurones” are descriptions of two different things⁴⁷,

⁴⁵ His chapter heading is apt and suggestive: “The Invention of Mind” – for Rorty, the mind is a post-Cartesian invention.

⁴⁶ Ryle (1960) uses this division of the world into two sorts of “stuff” to argue that Descartes had, fundamentally, perpetrated a category mistake, and rejects Cartesian dualism on those grounds.

⁴⁷ Cf. Ayer (1998:478-489) for an instance of such a position.

rather than two ways of describing the same thing.⁴⁸ He explicitly refers to Nagel's "What is it Like to be a Bat?" and questions whether speculations or thought experiments like this one can prove the non-physicality, and implied non-accessibility, of the mental in any way.⁴⁹ Rorty makes the salient point that the neo-dualists end up talking, not about how people feel, but about feelings as entities that can exist in veritable independence from their

⁴⁸ Rorty (1990:70) makes the assumption that the mind and body are made of two different kinds of "stuff" explicit in his chapter entitled "The Antipodeans," beings who refer to neural activity with the same ease and to the same effect as sensation and feeling are normally referred to by human beings. To the Antipodeans, having mental activity that can be regarded and named as wholly divorced from neuronal activity is as inconceivable as the reverse situation would be for dualists.

⁴⁹ Essentially Nagel (1982:391) is very sceptical about our ability to say anything meaningful about consciousness (about being). He is especially wary of attempts to describe consciousness in material terms and presents a thought experiment to illustrate his point, namely, he explores the possibility of knowing what it would be like to be a bat.

Nagel declares that most reductionist attempts to reduce mental phenomena to a variant of materialism fail, because they do not appreciate the distinctive difference between the mind-body problem and other problems that have successfully lent themselves to reduction: consciousness. In an attempt to rectify the situation Nagel proceeds to discuss consciousness, even though "it is difficult to say in general what provides evidence of it" (392). These qualms aside, Nagel lights upon the distinguishing characteristic of consciousness, no matter in what form: for an organism to be conscious, there must be something it is like to be that organism, something it is like *for* that organism "to be" (392). Any analysis of mental phenomena would, according to Nagel, need to take into account this "subjective character" of experience. The problem being that for reduction to be successful, "phenomenological features of the mind" must be given a physical account, which seems impossible, given that they are subjective – they cannot be separated from their single point of view.

Nagel illustrates his point with a thought experiment: can we know what it is like to be a bat? The answer, predictably, is no. The argument seems to be that, seeing that conscious experience is necessarily subjective, it is not possible to explain it in objective terms. We cannot know what it is like for a bat to be a bat. But, so Nagel argues, ratifiable knowledge needs to be objective. Mental states need to be known through the observation of physiological processes, and the species-specific viewpoint must be eliminated. Essentially, if mental processes are physical processes, Nagel argues that there must be something that it is like to be a physical process. He cannot accept this as a possibility, and as a result he concludes that, in order for us to have any legitimate knowledge of the mental, we need to develop an "objective phenomenology," which does not rely on empathy or imagination. Through this method we should be able to describe subjective experiences to those who cannot experience them. See Hofstadter (1981:403-414) for a convincing critique of the assumptions embedded in Nagel's thought experiment. In essence he dismisses Nagel's as an attempt to "subjectively know what it is objectively like" to be (409). Hofstadter believes that Nagel's fundamental error is the assumption that justifiable knowledge is objective. He emphasises the subjective element involved in knowledge and highlights the role that language plays in our ability to exchange ideas and experiences - knowledge. (See chapter 5 for a lengthy discussion of the issues involved in this debate, as they are fundamental to how one would construe the self and our ability to say anything meaningful about it.)

particular instantiations (30).⁵⁰ Rorty believes that we intuitively identify the phenomenal with the immaterial, as the result of our particular tradition, and hence give them a “non-spatio-temporal habitation” (31).

Rather than crediting these “intuitions”, Rorty proposes that intuitions are nothing but familiarity with a language-game, in the Wittgensteinian vein. To discover the source of these intuitions, we need to turn to the history in which this philosophical language game developed. Briefly, Rorty believes that philosophy had its origin when it became necessary, in antiquity, for something general to be said “about our knowledge of universals” (1980:38). This question was answered in terms of the metaphor of knowing general truths by internalising universals, just as the eye of the body knows particulars by internalising their individual colours and shapes. The answer, according to Rorty, that Western philosophy eventually comes up with for the question as to what makes man unique, is that man has an immaterial soul capable of contemplating universals.

In what seems to contradict his earlier misgivings about the universality of distinguishing between “two sorts of ‘stuff’” that constitutes the human being (1990:17) Rorty asserts that throughout history of Western philosophy human beings have been accorded two sides: the grossly material and what Rorty calls our “Glassy Essence”⁵¹, the finer part of our being through which we understand universals. The alternative, materialism, is disfavoured by tradition because of its prosaic implications for the soul. As Rorty so vividly puts it: “To suggest that the mind is the brain is to suggest that we secrete theorems and symphonies as our spleen secretes dark humors” (44). He concedes that some kind of conception of a “glassy essence” to have present in ancient and subsequent philosophy, and contrary to his earlier protestations, it would be safe to assert that Rorty *does* believe a vague kind of dualism to have been present throughout the history of Western

⁵⁰ A particularly problematic case is that of pain, which neo-dualism suggests cannot be a physical entity, “because it is phenomenal” (Rorty 1990:30-31). Later we will discuss Dennett’s (justified) dismissal of “phenomenality”.

⁵¹ Rorty defines this term as encompassing “all things which corpses do not have and which are distinctively human” (1990:44).

philosophy. But he insists that recent philosophy has lumped together the traditional idea of our “glassy essence” with the very different post-Cartesian notions of “consciousness” or “awareness” (45). And of importance to us is that Rorty believes modern conceptions of self to stem from these *post-Cartesian* notions. It seems that, although Rorty is sceptical of the claim that the mind-body distinction is intuitive, he does believe that there had been a separation between mind and body throughout Western philosophy, at least, and that this distinction means different things and is argued for through different philosophical theories both before and after Descartes (62).

For Rorty the mind-body distinction in its modern manifestation had its origin with Descartes, when he changed the conception of *mind-as-reason*⁵² to *mind-as-inner-arena* (61). Rorty sees the Cartesian change from mind-as-reason to mind-as-inner-arena primarily as the triumph of “the quest for certainty over the quest for wisdom” (*ibid.*) In other words, the quest to establish certain philosophical knowledge, modelled on the newly developed practises of the physical sciences. The task of the modern philosopher became that of obtaining certain knowledge through mathematical rigour, rather than to help people attain peace of mind. “Science, rather than living, became philosophy’s subject, and epistemology its centre”(61). With time, conceptions of this glassy essence changed from soul as a vaporous breath that permeates the body and survives the death of the body, to that of the rational mind, which elevates human beings above the brutes.

According to Rorty, it was Descartes’ work that allowed for the idea of *mind* with an existence *separate* from that of the body.⁵³ His analysis enabled philosophy to draw a line between “the cramps in one’s stomach and the associated feeling in one’s mind”(62). Descartes made use of the only

⁵² Reason here is coupled with what Rorty calls the “hylomorphic epistemology”, which, in keeping with Aristotle, thought of grasping universals as instancing in the mind what the frog instances in its flesh (Rorty 1990: 45; 62). He contrasts the hylomorphic model of mind with the Cartesian representative model (46). He sums up this position (which we endorse) as follows: “But if we see that the two models – the hylomorphic and the representative – are equally optional, perhaps we can see the inferences to mind-body dualism which stem from each as just as optional” (46).

⁵³ Toulmin (1990) takes up a similar position (37-40).

criterion that Rorty believes makes this distinction possible: indubitability⁵⁴. From there then the practice in contemporary philosophy to speak of “pains” and “feels” as a thing independent from the body, “situated” in a non-spatial, non-extended substance. The reason that the phenomenal and the immaterial are lumped together as the mental is because Descartes bridged the gap between the two with his notion of “incorrigibly known” (69). But, Rorty goes on to argue that the dualism of contemporary (analytical) philosophy in its turn is very different from Cartesian dualism⁵⁵ (63).

According to Rorty contemporary inheritors of the Cartesian distinction between mind and matter have lost touch with the seventeenth-century guise of “substance,” (thanks in part to Kant). Such philosophers interpret Descartes’ distinction between mind and body as distinct entities “as a recognition of the difference between parts of persons and the states of those parts ... on the one hand and certain states of the whole person on the other...” (66). Mental entities become subject to a stream of consciousness and a body – states of persons rather than “bits of ghostly stuff” (*ibid.*). The contemporary mind becomes non-spatial in the sense that the states of a person have a kind of adjectival status. The mind-body problem seems to disappear in this interpretation, as Rorty flippantly explains: “...few people are worried by an ontological gap between what is signified by names and what is signified by adjectives” (*ibid.*). While he does believe this solution to the mind-body problem fares well in terms of explaining beliefs and desires, he maintains that some difficulty is encountered when trying to explain pain and “raw feels.” While it seems (to Rorty at least) relatively unproblematic to accord non-spatiality to states (beliefs, desires, etc.), thoughts and mental images tend to be thought of as things, with a separate existence from the body. The ancients saw the “universal-grasping” substance as existing separately from the body, while contemporary dualists see *event-like* mental

⁵⁴ Rorty’s position compares well with Popper’s contention (discussed at the beginning of this chapter) that the emphasis on ultimate explanation is what sets Cartesian theory apart from its predecessors.

⁵⁵ Rorty (1990:67) distinguishes between four kinds of dualism: that between a person and his ghost, that between a person and his Aristotelian passive intellect, that

happenings as separately existing (67). It is clear from Rorty's discussion that he believes that neither the pre- nor the post-Cartesians shared Descartes' conception of the mind as "thinking substance," but that the trend that led to modern conceptions of mind was put in motion by Cartesian dualism. Rorty declares Descartes' only improvement on the idea of the intangible man to be his stripping it of its humanoid form (68).

For the purposes of this discussion, the controversy, which initially seemed to centre round whether the idea of the self, as such, existed prior to the Enlightenment, now seems in actual fact to centre around how a self-like entity was perceived prior to and subsequent to the Enlightenment. The significance of the Enlightenment lies, in its attempts to apply the newly formulated principles of the physical sciences to all aspects and subject matter of inquiry. All the theorists cited in this debate seem to agree to the extent that they believe conceptions of the self to have taken a radical and significant turn as a direct result of seventeenth century-theory.⁵⁶ A turn that has direct implications for current debates raging on self/mind and especially our current discussion on issues regarding the self, mind, consciousness, etc. In order to develop this argument we need to take an extensive look at Descartes' theories and subsequent developments and critiques thereof.

4. Some modern metamorphoses of self

Modern theorising on the self saw the emergence of two major opposing positions: *rationalism*, which saw the self, or more accurately perhaps, the subject, as disembodied thought or pure à priori mind, and *empiricism* which insisted that the subject was reliant on experience, and denied the possibility of à priori knowledge.

between a Cartesian *res cogitans* and *res extensa* and finally, the contemporary dualism, which allows for mental entities, *without the soul*.

⁵⁶ Toulmin (1990) proposes the interesting, and convincing, argument that this seventeenth century development was, in effect, a counter-movement to the gains that the Renaissance had made over medieval theories on the nature of man and the world. Refer to *Cosmopolis* (1990:5-80) for his extensively researched account of the origins and project of seventeenth century philosophy.

The dispute between the rationalist and empiricist positions basically boils down to whether reason or experience ultimately justifies belief - and by implication a dispute over how knowledge is acquired. Rationalist theory holds that the mind has innate ideas and that these innate ideas form the basis on which reason can form justifiable beliefs about the world. Descartes, for example, sees beliefs deduced through reason as the only possible justifiable beliefs and regards beliefs gained through experience as deceptive and untrustworthy. Reason here can be characterised as the mind's ability to discern the logical relations between ideas, in complete independence from experience (Radcliffe 2000: 30).

4.1 Intimations of Rationality

The self enters the modern philosophical era in the guise of Descartes' disembodied *cogito*. As legend has it, writing in the climate of the scientific revolution of the seventeenth century, Descartes sat himself down in a large Dutch oven, and proceeded to doubt everything that could possibly be doubted.⁵⁷ His aim was to attain the one sure premise on which knowledge can be grounded. Descartes was disillusioned with philosophy, which, despite being pursued by some of the most distinguished scholars in history, still, in his view, had not managed to reach consensus on any of the great philosophical questions. He declares his resolve to establish philosophy on a firm scientific basis as follows: "...when I considered the number of conflicting opinions touching a single matter that may be upheld by learned men, while

⁵⁷ Refer to Toulmin (1990: 152-161), where Toulmin discusses what he calls the "Standard Account" of the scientific revolution of the seventeenth century. According to the standard account, stagnancy and dogmatism of medieval thought came to an end when science took a rational turn. In the early seventeenth century Galileo proposed a science that was grounded on experimental observation, rather than subject to the authority of traditional philosophical speculation. In contrast to Aristotle, the Galilean world-view conceived of nature as mechanical and subject to mathematical and geometrical laws. And Newton's groundbreaking, *Mathematical principles of Natural Philosophy* would appear in 1687, perpetuating the work of Galileo.

Toulmin argues that the early seventeenth century was not only characterised by a scientific revolution, but also by the turmoil of the Thirty Years War and a backlash against the Renaissance, which would influence in many ways the trajectory of both philosophical and social developments of the Enlightenment (5-44).

there can be but one true, I reckoned as well-nigh false all that was probable" (1978:8).

In order to establish philosophical method on a firm foundation similar to that recently established for the physical sciences, Descartes set about finding certain knowledge: that which he could know to be "indubitable" seeing as "we may doubt in general of all things" (1978:75). In his search for concrete truth Descartes chose as his method that of rejecting all that he could reasonably believe to be knowledge based on opinion, and thus *not* indubitably true (1978:26). He defends his method of doubt as follows:

Now, although the utility of a doubt so general may not be manifest at first sight, it is nevertheless of the greatest, since it delivers us from all prejudice and affords the easiest pathway by which the mind may withdraw itself from the senses; and, finally, makes it impossible for us to doubt wherever we afterwards discover truth (75).

The reason that the mind needs to withdraw from the senses is that Descartes readily accepts the possibility that our senses may comprehensively deceive us in all that we perceive. Furthermore he poses the possibility of the existence of a *malignant demon*, who might present all that we perceive as real, while in actual fact it all is a dream⁵⁸ (84). If this is the case then all knowledge attained through the senses, or even the very idea of our possessing senses would be an illusion and is therefore useless in the acquisition of certain knowledge. Of course, if it were possible to acquire knowledge that could not be doubted that certainty would form the first principle on which other indubitable truths can be established.

As is attested in *Discourse on Method* (1978:27-32) and *Meditations on the First Philosophy* (1978:85-86), the only thing that could withstand this radical doubt was the very fact that Descartes doubted his own doubting. And negating the act of doubting would be self-contradictory, for what else is doubting but the act of thinking? Of course to be able to think one needed a

⁵⁸ Refer to Dennett (1991:1-10) for the modern day version of this thought experiment, featuring a brain in a vat and malignant neuroscientists.

thinker, which lead Descartes to the conclusion that there, necessarily, had to be a thinker, a *cogito*, whose existence could not be doubted without at the same time affirming its very existence.⁵⁹ It seemed, however, that this disembodied thinker could not be certain of the reality of anything other than itself, and seemed to be restricted to being a solitary and solipsistic entity.

After assuring himself that he exists, Descartes asks himself what exactly "he" is, especially in light of his radical doubt, in terms of which he could not even be sure of what he perceived through his senses (86). Descartes sifts through his preceding beliefs about his own "self" and attempts to reject all those beliefs that are not certain and indubitable:

What then did I formerly think I was? Undoubtedly I judged that I was a man. But what is a man? Shall I say a rational animal? Assuredly not; for it would be necessary forthwith to inquire into what is meant by animal, and what by rational, and thus from a single question, I should insensibly glide into others, and these more difficult than the first; nor do I now possess enough of leisure to warrant me in wasting my time amid subtleties of this sort. I prefer here to attend to the thoughts that sprung up of themselves in my mind, and were inspired by my own nature alone, when I applied myself to the consideration of what I was⁶⁰ (1978:86-87).

Upon retrospection Descartes discovers that he had previously vaguely thought of himself as consisting of body and soul; attributing to body all those aspects that a corpse would have (a countenance, arms, legs, ligaments etc.), and to soul aspects such as being able to walk, perceive, act, think etc. Descartes realises that he had never given much thought to what exactly soul was, always thinking of it in vague terms of something akin to "flame, wind or

⁵⁹ From there Descartes' famous declaration: "I think, hence I am", which he propounds to be the first principle of a scientifically sound philosophy (1978:27).

⁶⁰ From Hume onwards, criticism has been levelled at the very idea that thoughts can spring up "by themselves" in the mind, and are thus "inspired by [human nature] alone." This work throws its weight squarely behind those theorists who insist that knowledge (even knowledge of oneself) ante- cedes experience and that one cannot circumvent the messy business of insensibly gliding into a tangle of questions when discussing human being.

ether", permeating his body (87).⁶¹ Given his radical doubt and the possible existence of a deceiving and malignant being, Descartes is faced with the problem of which, if any, of his earlier beliefs about what it means to be a man, are still valid. He concludes that he cannot affirm any of the attributes of the body with certainty, and proceeds to a discussion on the attributes of the soul. Even though without a body it becomes impossible for the soul to have attributes such as nourishment, walking and perception, one attribute of the soul – thinking – proves to be indubitable, even in the absence of a body, since, as we have already mentioned, any act of doubting the possibility of thought, is in itself an act of thinking. Descartes declares that he has found the one thing that properly belongs to himself – he is certain that he exists, and that he exists as often as he thinks:

I am therefore, precisely speaking, only a thinking thing, that is, a mind (*mens sive animus*), understanding, or reason, - terms whose signification was before unknown to me (88).

Descartes is adamant that he "is" neither that which he had previously believed to belong to the body, nor that which he had seen as belonging to the soul (wind, flame, vapour, or breath). He is limited to being a "thinking thing" – a thing that doubts, understands [conceives], affirms, denies, wills, refuses, imagines, and perceives. Descartes initially expresses some doubt at his conclusion, saying that corporeal things still seem to be known with much greater distinctness than the "proper nature" he has persuaded himself that he possesses (90). But he doesn't take this doubt to mean that his supposition of what he is may be mistaken in any way - he declares his doubt to be the result of a wilful mind, not willing to submit to the restraints of truth (*ibid.*) He decides to "leave his mind" to its own devices, believing that it will inevitably succumb to the truth of his reasoning.

⁶¹ Descartes declares himself to have been absolutely certain of the attributes of the body, and, if pressed, he would have explained the body as all that can be comprised in a certain place and fill a certain space, to the exclusion of other bodies, and can be perceived through the senses. A body does not have the ability to move itself, but can be moved by another body. The abilities of self-motion and thought on the other hand, are not attributes of the body (1978:87).

Being certain of both the existence and the fundamental attribute of the soul, Descartes turns his attention to the corporeal world that his senses persist in presenting him with. Now that the incorporeal mind has been established as a certain truth, what other truths can be derived from this first principle? In his contemplation of a piece of wax (90-94) Descartes comes to the conclusion that bodies are not properly perceived through the senses or the faculty of the imagination, but only through the intellect, since they can only be perceived by being understood. From this realisation, Descartes concludes that nothing is more clearly apprehended or understood than one's own mind (94). In a further effort to accustom himself to this new opinion on mind, he undergoes a deliberate attempt to gain a more familiar and intimate knowledge of "himself" (95).

Since he had postulated the idea of the deceitful demon, who might cause him to perceive falsehoods through his senses, Descartes is not able to consider any other perception or imagining as being true. Unless, that is, he can prove the existence of God, seeing that God is the perfect being and a perfect being would not be purposefully deceitful (97). The existence of God is indispensable in Descartes' quest for indubitable knowledge.⁶²

Descartes argues that: "we may validly infer the existence of God from necessary existence being comprised in the concept we have of him" (170). According to this argument, the chief idea among others ideas in the mind is that of God – as an omniscient, all-powerful, and absolutely perfect being (170). Whereas ideas of other things contain ideas of possible and contingent existence, the idea of God contains the idea of absolutely necessary and eternal existence. Since we have this idea of an omniscient being, we should

⁶² Even though Popper (1977:179) speculates that Descartes might have added God to his argument to appease the church, in the light of what had happened to Galileo, it does not seem possible that Descartes could have constructed his philosophy without recall to the existence of God to guarantee certain knowledge, other than that of one's own mind. Toulmin rightly points that Descartes' theory was still in danger of offending the church, because it created the possibility of the world as a mechanical process, which might have been set up by God, but which can function on its own after the initial act of creation (1990:78-79). Judging from Descartes' own writing Popper's speculation seems unfounded, seeing that Descartes appears to have been sincere in his attempt to explain how we differ from animals, the answer that he came up with being that we possess a rational soul which is both immaterial and immortal.

enquire from where we could have acquired such a unique idea. Upon some consideration Descartes concludes that an idea of the perfect being cannot have its cause in anything less perfect than the perfect being itself, since the more perfect cannot arise from the less perfect. Every idea that we have, must have an original, which in itself possesses those perfections that we perceive. Since we are not at all perfect, omniscient or all-powerful, and since our less than perfect nature cannot give rise to the idea of something much more perfect than itself, it follows that we must have acquired our idea of God from an existent God (172-173).

He sums his argument up as follows:

But as we know that God alone is the true cause of all that is or can be, we will doubtless follow the best way of philosophising, if, from the knowledge we have of God himself, we pass to the explication of the things he has created, and essay to deduce it from the notions that are naturally in our minds, for we will thus obtain the most perfect science, that is, the knowledge of effects through their causes. But that we may be able to make this attempt with sufficient security from error, we must use the precaution to bear in mind as much as possible that God, who is the author of things, is infinite, while we are wholly finite (174-175).

Hence, he concludes, proper philosophising will entail examining the efficient causes of things, considered from *natural light*, (our faculty of reason), which we receive from God and which is therefore immune to deception. All that we perceive through our faculty of reason, and hence perceive clearly, is true, and consequently we are delivered from doubt (176-177). Descartes also distinguishes between two modes of thinking in us: the perception of the understanding and the action of the will (177), which is what distinguishes us from automata or animals.

The Cartesian reasoning mind is what makes us essentially human through endowing us with the capacity to move our bodies and other physical objects through willed and reasoned action. Descartes argues that if one is presented with intricately constructed automata (in the manner of the then *en vogue* hydraulic robots in the French Royal Gardens he would have been

familiar with)⁶³ one would, in the case of animal automata, be unable to tell the difference between sufficiently complex robots and real animals. With human automata, on the other hand, there would be certain dead giveaways: first, automata would never have the capacity for language sophisticated enough to convince onlookers of their state of minds, and second, no matter how skilled automata might be in the execution of the tasks that they were designed for, they would never be able to demonstrate having acted from knowledge, rather than in accordance with their design (1978: 44-45).⁶⁴ Man's superiority over animals and automata in this regard is not due to superior design on the part of man, or the lack of ability (i.e. the necessary organs etc.) on the part of animals. Descartes explains this discrepancy by declaring that animals and automata lack one important thing: reason. Descartes explicitly refers to man as possessing a "rational soul"⁶⁵ and to animals as being devoid of "mind" and as possessing a soul of a different nature from that of man (1978: 46)⁶⁶. Descartes' theories on the workings of the mind resulted in his mind-body duality, which postulates a disembodied subject, uninfluenced by contingent aspects of its corporeal body⁶⁷.

⁶³ See Flanagan (1991:1).

⁶⁴ This calls to mind Alan Turing's "Turing Test" as the best way of judging the "intelligence" in a machine. If the human mind had been simulated adequately on the machine the machine would, by way of typed responses to questions be able to convince its interrogator that he/she is corresponding with a human being. Turing leaves it open whether or not such a machine, if successful, should be considered conscious (Gregory 1998b:784).

⁶⁵ Note that the terms soul/mind, as with many theorists are also used interchangeably by Descartes.

⁶⁶ Most notably, the souls of brutes lack immortality (1978:46).

⁶⁷ One may question this extreme dualist position which is usually ascribed to Descartes in the light of statements of his, such as the following: "...[health]...is without doubt, of all the blessings of this life, the first and fundamental one; for the mind is so intimately dependent upon the condition and relation of the organs of the body, that if any means can ever be found to render men wiser and more ingenious than hitherto, I believe that it is medicine that they must be sought for" (1978:49). Descartes makes it clear, though, that the *essence* of man is not at all influenced by any aspects of the body: "And although I may, or as I shall shortly say, although I certainly do possess a body with which I am very closely conjoined; nevertheless, because, on the one hand, I have a clear and distinct idea of myself, in as far as I am only a thinking and an unextended thing, and as, on the other hand, I possess a distinct idea of body, only in as far it is an extended and an unthinking thing, it is certain that I [that is my mind, by which I am when I am] am entirely and truly distinct from my body, and may exist without it" (1978: 132-133).

Descartes argues that he can distinctly perceive mind and body as two separate things and he takes this to be sufficient evidence that they are substantially different from one another and that they have been made (by God) to exist separately from one another. He knows with certainty that he exists as a mind, while he has no reason to believe that his nature necessitates anything beyond being a thinking thing. In his certainty that his mind is distinct from his body, he concludes that his mind⁶⁸ can exist without his body (132-133).⁶⁹

One of the important advantages of Descartes' dualism is that it enables him to account for his belief that nature is mechanistic, while man is free to will (see Levin 1992:28). From this then Descartes' conclusion that all organisms, save human beings, are automata. The human body itself is an automaton, except with regard to its ability to have voluntary movement, which is made possible by the immaterial human mind.⁷⁰ For all his musings on possible ways that mind and body can co-exist, he never proposes a satisfactory explanation of how mind and body, as two distinct substances, can interact. (Descartes' speculation that such an interaction occurs in the pineal gland has long since been discredited.)⁷¹ It was this difficulty that lead to the transformation of Cartesianism by subsequent theorists.

Descartes' approach was quintessentially essentialist in Popper's sense of the word. Popper contends that his ideas rest upon an intuitive idea as to what the essence of man is, as is attested to by his method of radical doubt (1977:177). Popper also notes Descartes' "peculiar form of a mechanistic theory of causality," i.e. the idea that all causation in the physical world is caused by a mechanistic push (ibid.). Descartes applies this principle

⁶⁸ Descartes does not believe the mind and the brain to be the same thing, while he explicitly equates mind and soul (1978:139;140-141;218).

⁶⁹ Descartes does note that mind and body may interact, and even be, to an extent, interdependent, and he concedes that nature does "teach him" that he has a body (1978:134-135). But he insists that he cannot draw any conclusions with regard to external objects, without consideration by the mind and that the mind alone can discern the truth in his perceptions: "...for it is, as appears to me, the office of the mind alone, and not of the composite whole of the mind and body, to discern truth in those matters" (136).

⁷⁰ See footnote 36.

⁷¹ See Dennett (1991: 33-39).

of mechanistic causation to the mind as well, even though he believes the mind to be composed of a different substance altogether. This world-view is summed up with an apt analogy by Owen Flanagan: human bodies act upon the world in response to stimuli, in much the same mechanistic way as the life-size, hydraulically-controlled robots that Descartes encountered in the French Royal Gardens (Flanagan 1991:1).

In keeping with his distinction between ultimate and conjectural explanation,⁷² Popper declares that it is only in terms of such an attempt at ultimate explanation that the difficulty of their interacting would arise, because such an ultimate explanation is derived from the intuited essences of the mind and body with their apparently dissimilar constituting substances. In terms of conjectural explanation there should be no reason to pre-empt the possibility of their interaction (1977:182). Popper believes that it is in the attempt to combine an incorporeal soul/mind and the mechanistic notion of physical causation that Descartes encountered unnecessary difficulties and caused a shift in the mind-body problem, which subsequently lead to a mind-body parallelism and later to the identity thesis (177).

Popper's criticism hints at a certain arbitrariness on the part of Descartes in his mind/body division. Such a view, although unfounded, seems to be not all that uncommon. However, as Solomon and Higgins (1996:185) point out, Descartes' move is neither arbitrary, nor isolated, and could even be thought to be inevitable in the context of the history of philosophy:

We should not suppose, as is often charged, that Descartes made some sort of stupid mistake, arbitrarily marking off the mind from the body as different "substances" and then finding himself unclear about how to get them together again. The dualism of the mind and body was the product of several centuries of intellectual development, the progress of science and the newfound respect for individual autonomy. Distinguishing the mind and the body provided a realm for science, concerned with the physical world, to proceed unhampered by religion or moral concerns associated with

⁷² See above

the peculiarities of the human mind, human freedom, the human ability to “transcend” physical reality, and so on. The distinction also provided a realm for religion and human freedom and responsibility that would not be threatened by science. If the world from Aristotle to Aquinas had been largely defined by a single set of “natural laws,” whether provided by God or by nature, the new, modern world would have to juggle two sets of concerns, one for bodies, one for the mind (one for the facts, one for the values). From Descartes to Sartre, getting these two together would not be nearly as important as keeping them safely apart.

The problem of the mode of interaction between body and mind would dog Cartesian theory and lead to much criticism and attempts at modification.⁷³ The Cartesian self is “a substance whose whole essence or nature consist[s] only in thinking and, which, that it may exist has need of no place, nor is dependent on any material thing...” (1978:27). This thinking substance, or pure thought, detached and unaffected by any material substance hardly lends itself to elaborate discussion. It does not develop nor evolve, it remains pristine, untouched by the contingencies of life and remains independent from any particular body. Descartes’ cogito comes into the world with knowledge already imprinted on it and by means of reason we can bring these innate ideas to consciousness. One of the main tenets of the scientific methodology that Descartes aims to employ is, of course, to postulate universally valid knowledge. The assumption of universality makes it unproblematic for him to start with his own existence as paradigm example. And since he has found himself to be a thinking substance in essence, that attribute is generalised and applied to all people.

⁷³ A curious spin off, for instance, is the theory of occasionalism, a kind of psychophysiological parallelism. The occasionalists used Descartes’ own assertion that God is a perfect being and would therefore not deceive us to conclude that all causation is miraculous and the causation between body and mind was the result of the miraculous intervention of God (Popper 1977: 182). Later forms of parallelism – be they Spinozean or Leibnizean – would drop the call to miracles or God, but would still postulate a parallel functioning of mind and body, with no form of facilitation between the two.

Descartes' theories have received much criticism from many quarters, and have become, rightly or wrongly, the embodiment of all that is to be viewed with suspicion, or to be lauded – depending on one's position – in the Enlightenment. Ryle rightly attributes the furore raging around the Cartesian doctrine to the prevalence of the theory right up to the present day. Descartes' theory on mind has become the touchstone for most subsequent theories on the mind and related matters, and its influence on those theories – whether they are in agreement or in absolute opposition to Descartes' principles – cannot be overemphasised. Ryle (1960:11) believes that Descartes' legacy has so powerful a hold on "latter-day" (i.e. mid-twentieth century) theory that he describes it as the "Official Doctrine." Ryle's position is mirrored by that of Flanagan: "Descartes' theory remains the single most influential framework for discussing the philosophy of psychology and mind" (Flanagan 1991: 1). For many the difficulties with the theory are generally considered to be minor theoretical difficulties that can be overcome with minor modifications.

4.2 Descartes criticised

A telling critique of Descartes' theory of mind is that of Gilbert Ryle. He introduces his critique as follows. Ryle attempts to show that the *central principles* of Descartes' theory are unsound (his work merits a much more lengthy discussion than can be accorded to it here):

For certain purposes it is necessary to determine the logical cross-bearings of the concepts which we know quite well how to apply. The attempt to perform this operation upon the concepts of the powers, operations and states of mind has always been a big part of the task of philosophers. Theories of knowledge, logic, ethics, political theory and aesthetics are the products of their inquiries in this field ... but ... during the three centuries of the epoch of natural science, the logical categories in terms of which the concepts of mental powers and operations have been co-ordinated have been wrongly selected. Descartes left as one of his main philosophical legacies a myth which continues to distort the continental geography of the subject (1960:8).

In keeping with the tradition of language philosophy, Ryle criticises Descartes on the grounds that he has committed a category mistake.⁷⁴ Ryle argues that Descartes perpetrated such a mistake in that he performed certain operation with concepts of mind that are logically “improper” to apply to mental concepts and hence are “breaches of logical rules” (8). And, as noted earlier, he argues that the central principles of the Cartesian doctrine are unsound. Ryle takes the central principles to be the doctrine that human bodies are in space and subject to the mechanical laws that govern all physical bodies, while minds are not in space and thus not subject to its laws (11).

Ryle notes the Cartesian bifurcation where mind and body are said to occupy two different “worlds,” the physical and the mental (12). These two worlds are metaphorically spoken of as “external” and “internal,” although, as Ryle notes, strictly speaking, minds cannot be inside anything, seeing that they are not spatial at all⁷⁵. Even when the “inner” “outer” division is treated as metaphorical, difficulties still abound in trying to explain how the mind and body influence one another.⁷⁶ But Ryle’s criticism does not stop there. He claims that there is a deeper, philosophical assumption that underlies the Cartesian theory and gives rise to even more theoretical difficulties. Descartes makes the fundamental assumption that there are two different kinds of existence; two kinds of status when it comes to existing (13). As Ryle puts it: “...some existing is physical existing, other existing is mental existing” (ibid). Physical existing necessarily takes place in space and time, and consists of matter, while mental existing is necessarily in time, but not in space, and

⁷⁴ Ryle explains the perpetration of a category mistake in the following manner: “The logical type or category to which a concept belongs is the set of ways in which it is logically legitimate to operate with it...certain sorts of operations with the concepts of mental powers and processes [applied by Descartes] are breaches of the logical rules. I try to use *reductio ad absurdum* arguments both to disallow operations implicitly implied by the Cartesian myth and to indicate to what logical types the concepts under investigation are to be allocated” (1960:8).

⁷⁵ Here Ryle levels criticism against some Cartesian theorists who forget the metaphoric nature of this division and speak of the mind as if were located in the skull (1960:12)

⁷⁶ Descartes clearly construed the two as influencing one another, and did not adhere to parallelism: “...the mind is [...] intimately dependent on upon the condition and relation of the organs of the body...(1978:49).

exists as consciousness. Furthermore physical objects are mechanically connected and subject to the laws of cause and effect, while mental fields are insular, impenetrable to one another. Despite this impenetrability to other minds, a person has direct knowledge of the workings of his/her mind and has privileged access to his/her mind through introspection⁷⁷. Ryle asserts that a necessary consequence of this view is that it implicitly prescribes a special way in which our concepts of mental operations are to be construed (15).

Ryle famously calls the Cartesian doctrine “the dogma of the Ghost in the Machine” (ibid.). He claims that the entire doctrine is false in principle, because it rests on a category mistake, and he calls the doctrine a “philosopher’s myth”. In Ryle’s words: “It represents the facts of mental life as if they belonged to one logical type or category (or range of types or categories) when they actually belong to another”(16).⁷⁸ And Ryle sets about rectifying the “logic of mental-conduct concepts”. It is nearly impossible to sum Ryle’s argument up more succinctly, or more eloquently than he does in the following passage:

My destructive purpose is to show that a family of radical category-mistakes is the source of the double-life theory. The representation of a person as a ghost mysteriously enclosed in a machine derives from this argument. Because, as is true, a person’s thinking, feeling, and purposive doing cannot be described fully in the idioms of physics, chemistry, and physiology, therefore they must be described as counterpart idioms. As the human body is a complex organised unit, though one made of a different sort of stuff with a different sort of structure. Or, again, as the human

⁷⁷ Ryle suggests that this view has remained essentially unchanged by the theories of Freud, which Ryle sums up as showing “that there exist channels tributary to this stream [of consciousness], which run hidden from their owner” (1960: 14). According to Ryle the adherents to the official doctrine, he does not specify who he has in mind, insist that under normal circumstances a person must be directly and authentically aware of the state and workings of his/her own mind. In the following chapter we shall discuss the importance of Freud’s theories on conceptions of the mind, and the subject, and how his insights make such a position untenable.

⁷⁸ Ryle does not offer a definition of a category mistake, but illustrates his understanding of the concept with a series of illustrations (1960: 16-17). What his illustrations basically boil down to is to indicate that certain concepts are applied to “logical types” to which they do not belong.

body, like any other parcel of matter, is a field of causes and effects, so the mind must be another field of causes and effects, though not, (heaven be praised), mechanical causes and effects (1960:18).

Ryle traces the origin of the Cartesian category-mistake to Galileo and his methods of scientific discovery, which were aimed at providing a mechanical theory applicable to all objects in space, and proposes that Descartes might have found himself embroiled in a conflict of motives. He wanted to establish a philosophical methodology based on the methodology of the physical sciences, yet he could not accept the inevitable conclusion implied by such an assumption, namely that human nature is just a variety of the mechanical clockwork of the universe, and that the mental is subject to the same laws of cause and effect as physical bodies. In order to avoid this conclusion, Descartes construed mental processes to be non-mechanical and non-spacial processes. And, since mechanical laws explain the movements of bodies in space, they cannot be applicable to the non-spatial workings of the mind. These two realms of existence, the physical and the mental, are then subject to different kinds of causation, and the human mind is not subject to the mechanical laws of the physical universe.

Ryle insists that, although construed to be radically different, the physical and mental were still considered within the common framework of the categories of "thing", "stuff", "attribute", "state", "process", "change", "cause", and "effect". Minds were subject to causes, effects, states, etc. different from those of bodies. This assumption of a mental realm is what then leads to the central theoretical difficulty in explaining how the minds and bodies can influence one another. Ryle, as a language theorist, summarises his criticism as follows: "Still unwittingly adhering to the grammar of mechanics, he tried to avert disaster by describing minds in what was merely an obverse vocabulary" (20). Mind was explained in the negatives of descriptions given to bodies. Following this line of reasoning, it seems fair to infer that, if physical bodies are subject to mechanical laws, minds, since they do not belong to the same category as bodies, must be subject to different, non-mechanical, laws. And, in order to allow for free will, non-mechanical laws cannot be deterministic in

the sense that mechanical laws are. Essentially, Ryle believes that the problem of free will arose from the question of how to reconcile the hypothesis that minds and bodies belong to the same category of mechanics, with the idea that “higher-order human conduct” is different from that of machines (20).

Ryle argues that the seeming contrast of mind and matter is illegitimate - they are not polar opposites, because they are not of the same logical type. He also argues that both idealism and materialism are answers to an “improper question”, and that mind and matter cannot be reduced to one another, because such a move would presuppose the legitimacy of their disjunction in the first place (32).

Ryle's argument up to this point is convincing. What remains unconvincing is his conclusion that: “It is perfectly proper to say in one tone of voice, that there exist minds and to say, in another logical tone of voice, that there exist bodies”(23), because “existence” is not a generic word and this is an example of two different senses of “exist”. His implication seems to be that mind and matter belong to different categories, and that what needs to be done is to allocate the mental to its proper category and then to apply concepts to it that are logically acceptable to that category. His argument does not explain why mind and matter belong to different categories, or how these categories are to be determined. Although Ryle (rightly) declares the disjunction between mind and matter to be illegitimate, and therefore the reduction of the one to the other as non-sensical, he does not offer a viable alternative. We agree with the spirit of Ryle's criticism – the illegitimacy of the distinction between mind and matter – but we will try and substantiate this claim on different grounds.

Although Ryle contests that the three-hundred-years plus debate on mind principally waged between idealism and materialism to be based on the false assumption that mind and matter are indeed separable, some aspects of this philosophical debate will prove informative with regard to the origin of lingering perceptions on the subject. We will therefore proceed to discuss the influence that the Cartesian doctrine would have on philosophy of mind for the next three-hundred-odd years.

4.3 Intimations of Empiricism

Descartes' influential view was not the only one handed down from the Enlightenment. His rationalism already received contemporary critique in the form of empiricism.⁷⁹ The most well known of the empiricists, David Hume, is not very taken with the Cartesian idea of identity as a given and eventually concludes that identity is an illusion altogether. As with the Cartesian project Hume wanted to put the humanistic sciences on as firm a basis as the seventeenth century physicists had put the physical sciences on, by using a similar methodology when contemplating general truths about human existence.⁸⁰ Hume's famous scepticism (and here he is diametrically opposed to Descartes) lies in his belief that claims of reason claim more than is their due, and he sets about exposing the limitations of human reason.⁸¹ Hume's project was to establish how we come to have knowledge, based on observation and reasonable inference, rather than through reason alone. In keeping with his status as a staunch empiricist, Hume's approach was to study the way that the mind functions, avoiding any *à priori* ideas about the workings of the mind. By means of epistemological analysis, Hume sets about

⁷⁹ The Empiricist Locke, for instance, insisted on the primacy of experience to knowledge and described the mind as a kind of *tabula rasa*, which needs to be inscribed with knowledge gained through experience. This knowledge consists of abstractions from sensation. Locke muses on identity, also personal identity and so doing becomes one of the first modern day philosophers to explicitly raise the problematic nature of personal identity, of self (Levin 1992: 19).

Locke explains personal identity or the self as "the I that accompanies all consciousness" (Levin 1992:21). We can have identity in several senses that include the other three kinds of identity distinguished above. Self-consciousness necessarily accompanies consciousness, but our sense of identity is not disturbed by breaks in consciousness. This becomes possible through memory, as memory bridges the gaps between breaks in consciousness. My memory of my past consciousness as well as my experiencing the organisation of my body as enduring through time becomes the basis of personal identity. Locke becomes one of the first modern philosophers to find a link between a sense of self and the body with its sensations.

⁸⁰ While Levin (1992:28-29), for example, explicitly refers to Newton when discussing Hume's aim to apply the new scientific method of the seventeenth century to human studies, some theorists like Radcliffe (2000) implicate Bacon. Be that as it may, Hume envisioned humanistic studies based on the newly developed scientific method applied both observation and reasoning, and which aimed to be objective – a project seemingly no different from that of Descartes, but which would lead to radically different conclusions.

⁸¹ To quote Levin: "He wants to be reasonable rather than rational, and in the final analysis relies on sentiment and custom to validate a great deal...[and to] determine human action" (27).

piecing together the structure of human thought. He explores the origins of our beliefs and tries to evaluate their foundation in reason. Of this expected foundation in logic he finds none and comes to the conclusion that our beliefs are rooted in custom, habit and sentiment.⁸²

In order to establish how the mind functions and how knowledge is acquired Hume sets about examining “mental contents”. Hume calls what is usually understood under mental contents *perceptions*, which are each of them an entity in themselves and are essentially sensations – sounds, tastes, odours, pressures, etc. Humean perceptions, or objects of mind, are atomistic entities, which are further divided into *ideas of sensation* (which could be simple or complex and entail thinking and reasoning) and *impressions* (which are feelings and experiences). Moreover, the latter are divided into two kinds of impressions: impressions of *sensation* - those that are the result of the external world affecting the senses and; and impressions of *reflection* – memories and fantasies (Hume 1969/1739: 49-56). Hume himself had difficulty establishing a satisfying way of distinguishing between the two. He took recourse to the vividness and intensity of impressions by means of which one could supposedly distinguish between the more immediate impressions of sensation, and the less vivid impressions of reflection:

ALL the perceptions of the human mind resolve themselves into two distinct kinds, which I shall call IMPRESSIONS and IDEAS. The difference betwixt these consists in the degree of force and liveliness, with which they strike upon the mind, and make their way into our thought or consciousness (1969/1739: 49).

Hume, as a sceptic, wants to expose the limitations of reason. He believes that feeling and sentiment determine human action, and relegates reason to being a mere “slave to the passions”:

“...all our reasonings concerning causes and effects are deriv’d from nothing but custom; and that belief is more properly an

⁸² He is often criticised, though, for not accomplishing this goal and for mistaking many *à priori* assumptions for observations (Levin 1992:26; Radcliffe 2000:14).

act of the sensitive, than of the cognitive part of our nature” (Hume 1969/1739:234).

By distinguishing the contents of mind in terms of their “feel” (vividness), Hume already sets us up to think of experience in terms of its quality and not its cause (be it external, or through mental processes) (Radcliffe 2000: 8). He argues that all our ideas can be traced to impressions, and that one cannot entertain an idea in the absence of an accompanying impression. A person cannot have the idea or the concept of a smell or a colour without having experienced it. *Experience* tells us, Hume believes, that the mind *cannot* have any innate ideas. Hume’s expulsion of innate ideas does not preclude the possibility of having ideas of the imagination. But, in Hume’s world, imagination is limited by previous sensory experience. No matter how improbable one’s flights of fancy, one cannot imagine something that one has never come across in any context before. In short, one cannot conjure up an original idea without any mental contents, and mental contents are all subsequent to experience. Hume ultimately comes to the conclusion that *reason* plays a very small role in determining our actions, and leaves us with a conception of human nature that is almost as mechanistic as Descartes’ conception of nature independent of mind.

Hume’s empiricism leads to an epistemological schema in terms of which human knowledge consists of impressions of sensation, impressions of reflection and, ideas – these are the *only* sources of knowledge. Hume’s contention that all ideas occur in the wake of impressions, and thus experience, raises the question as to how *abstract* ideas (like those of identity) are possible. We do not experience abstract ideas through our senses, so how are they formulated? Hume must explain what prompts us to adopt beliefs whose scope goes beyond the experiences on which they are based. Hume is very aware of the necessity of accounting for this possibility, but whether he addresses the possibility of abstract concepts successfully debatable. His answer to this dilemma is that ideas are relational, by virtue of a uniting principle. A bond between ideas will cause associations to arise in the mind that will highlight possible connections between ideas. Hume proceeds to elaborate on the nature of this relational bond:

He begins by asserting that of all the kinds of “philosophical” relation (1969/1739:117) that can exist by means of association in the mind (there are seven in all), only the principle of *causation* can allow us to make inferences from present experience: “Here it then appears, that of those three relations, which depend not upon mere ideas, the only one, that can be trac’d beyond our senses, and inform us of existences and objects, which we do not see or feel, is *causation*” (122).

If the experience of apparent causation in the mind is the only possible origin of abstract concepts, how does this kind of causation arise? Hume insists that nothing exists, internally or externally, that cannot be considered either a cause or an effect at some time or another. At the same time he argues that it is very clear that there is no universally existing quality common to all of these instances, which allows them to be classified as causes or as effects (123). Hume proceeds to distinguish between *causality* and *correlation*. He asserts that our idea of causality has three component ideas: 1) *contiguity*: two experiences have to be conjoined in time and space in order for us to experience them as being causally connected; 2) *succession*: cause must seem to precede effect; 3) *necessary connection*: for two correlating experiences to be causal there needs to lie a necessity in their causal relation, a similar cause, under similar circumstances must produce a similar effect (Hume 1969/1739:123-126). So far so good, contiguity and succession are sufficient for us to establish a correlation between two experiences. But, according to the preceding schema, if we are looking for an instance of causality we need to bring necessity into the equation - a particular cause must *necessarily* cause a particular effect. Hume insists, however, that he cannot find any experiential examples that imply such a necessity. This leads him to conclude, famously and contentiously, that that causality is not an attribute of the world, and there is no reason to believe that events which we perceive as causally related will prove to be so again in the future (126-130). Yet, he argues, we are still able to make accurate judgements of cause and effect.

For Hume, we establish an apparently necessary connection between perceptions⁸³ through experience. But when we examine our external experience we only experience a set of impressions, followed by a further set of impressions, we do not find the impression of a connection between a set of impressions. This leads him to conclude that causality is in fact the result of a psychological process, where, because of experiencing certain events in conjunction with one another, we form the *habit* of thinking of the one in connection with the other, which leads us to posit a causal connection between the two. When we experience this anticipating of one event upon another, we experience an impression to which can be traced the idea of necessary connection, and causality becomes non other than an anticipation in the mind of cause and effect (Hume 1969/1739:126-135). Both memory and experience are what allow us to make these inferences:

There is no single phenomenon, even the most simple, which can be accounted for from the qualities of the objects, as they appear to us; or which we could foresee without the help of our memory and experience (Hume 1969/1739:117-118).

If the objects of mind (impressions) originate externally, or in memory and fantasy, and are related and ordered by means of association, we are left with a theory of the working of the mind that is not rational, but based on habit and custom which leads us to experience ideas as causally related (1969/1739:152).⁸⁴

Given Hume's conceptual schema, and his conclusion with regard to abstract concepts, a vexing question arises: how do we come to an idea of the self or mind as a subject that experiences these objects? If all ideas are traceable to impressions, there must either be an impression or a set of

⁸³ Hume uses perception as a generic term that denotes both impressions and ideas.

⁸⁴ Levin explicitly likens Hume's thesis on causality to Newton's Gravitation, where the association of ideas is the force that relates atomistic impressions, as gravitation might do in terms of physical bodies (1992:28). Economist Adam Smith, a contemporary and friend of Hume's, presents us with a gravitational force of his own, which is also analogous to Hume's causality and apparently influenced Hume's conception of it. Smith's "invisible hand" theory postulates the market and its laws to be operating along the lines of the same principles as the association of ideas in Hume (Levin1992: 29).

impressions from which the idea of self is derived (299), or, the self, as with other abstractions, might owe its existence to habit and custom.

As can be imagined, Hume does not balk at the idea that a concept of the self is based on custom rather than on a perceivable phenomenon. In short he concludes:

If any impression gives rise to the idea of self, that impression must continue invariably the same, thro' the whole course of our lives; since self is supposed to exist after that manner. But there is no impression constant and invariable...It cannot therefore be from any of these impressions that the idea of self is deriv'd; and consequently there is no such idea (Hume 1969/1739:300).

His argument runs as follows: Although our perceptions are in flux, we always experience them as being the impressions of someone, of a self, which means that we should have a continuous impression from which the idea of the self is derived. But, search as we might, such an impression is not forthcoming. Hume, as a sceptic, does not believe our experience of reality to be continuous, but rather to be atomistic. Given that abstract concepts can only be either impressions or relations of ideas, the self can only be known as one of the two. The self, as an impression (an object of experience) or as relational, or as unknowable, are the only options that Hume's epistemological schema allows as conception of the self. Hume dismisses the notion that the self is a relation of ideas off the cuff. He concedes that there is the logical possibility that an object "self" from which we could derive the idea of self might exist, but in keeping with his scepticism, he believes that it is impossible for us to know (300).

Having dismissed the possibility of a relational self, Hume, in terms of his epistemological scheme, is left with the possibility of having an impression of self. Hume declares that whenever he tries to enter or capture his "most intimate self" he inevitably comes upon an impression of some sort or another, whether it be heat or cold, pain or pleasure, etc. (300). He cannot catch himself without a perception of some sort. The absence of such perceptions only occurs when he cannot be sensible to himself, when he is

sleeping, for example. To quote Hume: "When my perceptions are removed for any time as by sound sleep; so long as I am insensible of myself, can I truly be said not to exist" (ibid). Hume declares that upon introspection that he always experiences a flow of impressions, but never an experiencer, as it were, of those impressions. This leads Hume to conclude that man (or the self) is nothing but a bundle or a collection of perceptions that are in rapid and perpetual flux.

Hume chooses as his metaphor of mind a theatre, where perceptions successively make their appearance and then pass off the stage – perceptions can also "mingle in a variety of postures and situations" (301). Hume warns us that the metaphor should not be stretched too far; mind is constituted only by a succession of perceptions. We have no notion where in the mind these scenes are represented nor of what they are made. It is only the existence of memory that allows us to experience these successive perceptions as continuous. Memory enables us to envision a chain of causes and effects, which, in the end, constitute our self or personal identity. Memory is not infallible, however, and the task falls on us to fill in the gaps in memory. In postulating the self as succession of perceptions, Hume has left us with the ability to experience perception, but not to experience the self that does the experiencing. We have the ability to create a sense of a continuity that does not exist externally to us, and which enables us to experience a sense of personal identity. But we are unable to know the self as such.

Hume proceeds to ask why there exists "so great a propensity to ascribe an identity" to successive perceptions and to ascribe to ourselves an invariable and uninterrupted existence throughout the course of our lives (Hume 1969/1739:301). In answer to this question Hume distinguishes between personal identity with regard to our thought and imagination; and with regard to our passion or concern that we take with ourselves. He proceeds to explain that it is by means of our imagination that we fill in the "gaps" in our experience, and find grounds to link a succession of objects experienced and thus render them, to our mind, an identical object through time. The identity that we ascribe to "the mind of man" stems from a similar operation of the imagination upon objects and hence our identical self is a fictitious one. In principle this ascription of objects of mind to be related and

belonging to a self identical through time occurs in the same manner as causality does in the case of external objects, from a kind of habitual association of ideas. "For from thence it evidently follows, that identity is nothing really belonging to these different perceptions, and uniting them together; but is merely a quality, which we attribute to them, because of the union of their ideas in the imagination, when we reflect upon them (Hume 1969/1739:307). Memory acquaints us with the succession of perceptions and is therefore the chief source of personal identity (309). Hume qualifies this statement by asserting that memory does not so much produce personal identity as discover it, by showing us the relation of cause and effect among different perceptions. In this way he leaves the way open for identity to be extended beyond the limitations of memory (310). Hume concludes his argument by declaring that identity is a grammatical, rather than a philosophical question: "All the disputes concerning the identity of connected objects are merely verbal, except so far as the relation of parts gives rise to some fiction or imaginary principle of union, as we have already observed" (310).

In the appendix to the *Treatise on Human Nature*, Hume expresses his own discomfort with his conclusions on personal identity and self, but declares: "I neither know how to correct my former opinions, nor how to render them consistent (1969/1739:675). Hume still holds that that one cannot perceive a self independent from perceptions, and that the composition of perceptions form the self (676). Hence: "...we have no notion of [mind], distinct from the particular perceptions" (677). What Hume finds problematic with his own theory is how to explain the principle according to which these perceptions are connected. We cannot perceive such connections, and Hume believes that we only *feel* a connection or determination of a thought to pass from one idea to another. He concludes that: "the thought alone finds personal identity when reflecting on the train of past perceptions, that compose a mind, the ideas of them are felt to be connected together and naturally introduce each other" (ibid.). Hume cannot find a satisfactory theory to explain the principle that connects our successive perceptions in thought, and his theory runs into two contradictory principles: "*that all our distinct perceptions are*

distinct existences, and that the mind never perceives any real connexion among distinct existences" (677).

In spite of his extreme scepticism, and his radical contribution to the debate on knowledge that had raged since the commencement of the Enlightenment, Hume, as Solomon and Higgins (1996:197) point out, readily accepted the dualism of body and mind. They believe (as does Ryle) that the dispute between rationalism and empiricism is better described as a "family feud", rather than an ideological debate. Both positions embodied a closely related sense of reason and other fundamental assumptions – the debate largely raged around innate ideas, the scope of reason and the method through which belief can be justified - finer detail, rather than fundamental theoretical differences. Both positions presented a united front, as it were, against dogma and superstition, which they believed could be countered with the universal human capacity for reason. Solomon and Higgins make the acute observation that modern philosophy is "not an extended debate about ontology, epistemology, and metaphysics...Irrationality was their true target. The enlightenment was not about the nature of knowledge so much as it was a defence of knowledge and inquiry ...The debate between reason and experience was a strategic technical distraction within the bounds of this cosmopolitan movement" (1996:199).

To the empiricist reason has its limits, and Hume advocates the importance of individual character – a good upbringing, cultivation of the virtues, a respect for traditions. And, although reason has its limits, our sentiments and our natural common sense, cultivated through our social traditions, have power and virtue to address the limits of reason, and have, to Hume's mind at least, been too long neglected in the overtly scientific atmosphere of modern philosophy (198).

Immanuel Kant, although an admirer of Hume, would not be able to reconcile himself with Hume's scepticism and aims to save the Enlightenment project by establishing the limits of what we can and cannot know. The work of Immanuel Kant will make up the last chapter in this "family feud." The Kantian self proves to be an amalgamation of both the rationalist and the empiricist point of view and only in the twentieth century, with the work of

Sigmund Freud, would the self take a radical turn and begin to shake off its Kantian trappings.

4.4 The self becomes transcendental

Hume believed that by empirically studying the mind he could achieve knowledge about the characteristics and the limits of human cognition. Kant, although an admirer of Hume, disagreed with his analysis of the limits of the mind. With his transcendental metaphysics Kant sets about establishing the conditions necessary for cognition, which need *not* rely on habit and custom and which can be justified objectively. His project would also be one in terms of which he aims to establish the limits of what can and cannot be known by the mind, and to establish how one can know judgements to be certain and universal. In keeping with the trend of the Enlightenment, Kant tries to reconcile the new scientific knowledge as exemplified by the mathematical world-view of Newton, with human freedom, and to explain the possibility of both. Science would be shown to be rationally justified, while moral law would be shown to be universal.

Kant inherited from his predecessors in the empiricist tradition a conception of human freedom seen as subordinated to the workings of a mechanistic universe. From his rationalist predecessors he inherited a picture of human freedom as based on rational insight into an objective realm inaccessible to the senses (Guyer 1995:2). One of Kant's invaluable contributions to Western thought would be his recognition of our own input into our perception of the world, presenting the human being as an active agent in cognition. Kant's point of departure in his inquiry into cognition is not to look for the self, but one the one hand to establish the conditions that are necessary for us to have coherent experience, and on the other to formulate his philosophy as a science with predicative and explanatory power. In short, his eventual conclusion would be that a self-consciousness must *necessarily* accompany every mental act, which follows that the self must exist. Subsequent to Kant neither science, nor morality could be thought of as the result of the passive reception on our part of objective truths or reality.

As we have seen in this chapter so far, causality has played an important part in the theories of self of both Descartes and Hume. Hume uses the concept of causality to illustrate that concepts that we do not empirically experience through the senses are validated on the grounds of habit, rather than on the grounds of being universally certain. Kant is uncomfortable with the Humean conclusion that causality is the result of habit and custom, rather than having an ontological basis. He sets about proving that the concept of causality is *à priori*, and therefore universal and certain (although, as we shall see, it can still be relegated to the category of *impure à priori* knowledge). This dispute reflects Kant's theoretical stance that there are, in the sphere of human cognition, structures that enable and ensure strictly *à priori* knowledge. Kant does not dispute the idea that "all our knowledge begins with experience," but argues that conceding that all our knowledge begins with experience, does not necessarily imply that all knowledge *arises* out of experience (Kant 1781/1990: 1-3). He distinguishes between *à priori* and *à posteriori* knowledge in the following manner:

By the term "knowledge *à priori*," therefore, we shall in the sequel understand, not such as is independent of this or that kind of experience, but such as is absolutely so of *all* experience. Opposed to this is empirical knowledge or that which is possible only *à posteriori*, that is, through experience. Knowledge *à priori* is either pure or impure. Pure knowledge *à priori* is that with which no empirical element is mixed up. For example, the proposition "Every change has a cause," is a proposition *à priori*, but impure, because change is a conception which can only be derived from experience (1781/1990:2).

Kant explains why he still considers this proposition to be *à priori* in the following manner:

Now, that in the sphere of human cognition, we have judgements which are necessary, and in the strictest sense universal, consequently pure *à priori*, it will be an easy matter to show ... If we cast our eyes upon the commonest operations of the understanding, the proposition "every change must have a cause," will amply serve our purpose. In the latter case, indeed, the

conception of cause so plainly involves the conception of a necessity of connection with an effect, and of a strict universality of the law, that the very notion of a cause would entirely disappear, were we to derive it, like Hume from a frequent association of what happens with that which precedes, and the habit thence originating of connecting representations – the necessity inherent in the judgments therefore merely subjective (1781/1990:3).

As a criterion for determining empirical from pure cognition Kant calls upon the ideas of necessity and universality. A proposition that contains the idea of necessity in its very conception is, inevitably, an *à priori* judgement. A judgement is absolutely or pure *à priori* when it derives from a proposition which equally contains the idea of necessity. As far as universality is concerned, empirical propositions are (as Hume had shown) never strictly or absolutely universal, but are assumed to be universal (i.e. comparatively universal) by means of induction. *À priori* judgements, on the other hand, are strictly and absolutely universal (2-3).

Kant, like Hume, is concerned with the conceptions that seem to extend our judgements beyond their experiential bounds. Here, in the transcendental⁸⁵ realm, where our knowledge does not arise from experience, we encounter Kant's conception of Reason. Reason takes as its subject matter those problems that cannot be addressed on the grounds of experience – problems such as God, freedom (of will) and immortality (5). The science by means of which these problems are usually approached is Metaphysics. But Kant describes metaphysics as being dogmatic from the outset, especially in that it does not even question whether reason is able to deal with such questions. How do we arrive at such *à priori* judgements, and how do we determine their validity? Some forms of pure *à priori* knowledge, like mathematics, have long been established and Kant argues that the kind of thinking that assumes this success can be translated onto all aspects or

⁸⁵ Kant explicitly defines his concept of the transcendental as follows: "I apply the term transcendental to all knowledge which is not so much occupied with objects as with the mode of our cognition of these objects, so far as this mode of cognition is possible *à priori*" (1781/1990:15).

pure knowledge, without allowing for the fact that such aspects may be of a different nature, is deceptive. He stresses the necessity to examine the origin of the cognitions on which we would build a metaphysics before we can accept them as principles, and he cautions against complacency that would disregard possible limits to reason (5). With his *Critique of Pure Reason* Kant sets about determining these limits, in other words, a critique of the faculty of reason.

In his critical assessment of Hume, Kant starts off by examining Hume's categories of judgement, where all knowledge is either empirically determined matters of fact, or logically determined relations of ideas. Abstract concepts can by definition not be empirically experienced and are therefore the result of the habitual relation (or synthesising) of certain ideas. Kant's assessment of judgement is much more intricate (1781/1990:1-15). He makes use of basic categories of judgement, namely, *à priori/à posteriori* and analytic/synthetic and as a result he presents us with four possible kinds of judgement: *analytic à priori*, *analytic à posteriori*, *synthetic à priori*, and *synthetic à posteriori*. Of these four kinds *analytic à posteriori* is by definition impossible – an analytical proposition does not rely on experience to be verified, its conclusion is already implied in its premises. Analytic judgements are by definition *à priori*, so the possibility of analytic *à priori* does not pose a problem. Similarly, the idea of synthetic *à posteriori* poses no problem, one cannot determine a synthetic state of affairs, without *à posteriori* experience.⁸⁶ What is at issue here for Kant and also for the purposes of our discussion is the possibility of the *synthetic à priori* (Hume does not consider it a possibility):

⁸⁶ For absolute clarity on what Kant means by synthetic and analytic judgements I quote at length from the *Critique of Pure Reason*: "Analytical judgements (affirmative) are therefore those in which the connection of the predicate with the subject is cogitated through identity; those in which this connection is cogitated without identity, are called synthetical judgements. The former may be called explicative, the latter augmentative judgements, because the former add in the predicate nothing to the conception of the subject, but only analyse it into its constituent conceptions, which were thought already in the subject; the latter add to our conceptions of the subject a predicate which was not contained in it, and which no analysis could ever have discovered therein" (1781/1990:7). Predictably, judgements of experience are always synthetical.

Upon such synthetical, that is augmentative propositions depends the whole aim of our speculative knowledge *à priori*; for although analytical judgements are indeed highly important and necessary, they are so only to arrive at that clearness of conceptions which is requisite for a sure and extended synthesis, and this alone is a real acquisition (1781/1990:9).

Kant believes that the theoretical sciences of reason – mathematics, physics, and metaphysics – consist of synthetic *à priori* judgements as principles, and rather than abandon these disciplines to habitual (and perhaps arbitrary) experiences of relation, Kant seeks to ground such knowledge as universally certain. For Kant, logically grounded knowledge of the world is only possible through *verifiable*, synthetic *à priori* judgements, and one of Kant's transcendental⁸⁷ conditions for *synthetic à priori* judgements is the existence of a continuous self.

Kant believed that we process experiences derived from external sources in one of two ways: aesthetic and categorical. Through his Transcendental Aesthetic⁸⁸ Kant determines that we invariably order our experiences of the world, our sensations, in terms of space and in time (1781/1990:21-43). Kant concludes that the objects of our experience are phenomena - "The undetermined object of an empirical intuition is called a phenomenon" (21). We experience the objects of empirical intuition invariably as being in space and time. Furthermore, Kant argues, while we can conceive of space and time devoid of objects, it is impossible for us to experience objects independent of spatial and temporal location. From this Kant concludes that space and time are logically *prior* to sensory representation (therefore, *à priori* categories of understanding), where space is the intuition

⁸⁷ Kant circumscribes his use of the term transcendental as follows: "I apply the term transcendental to all knowledge which is not so much occupied with objects as with our mode of cognition of these objects, so far as this mode of cognition is possible *à priori*" (1781/1990:15).

⁸⁸ Defined as "the science of all the principles of sensibility *à priori*," while the principle of pure thought is called transcendental logic (1781/1990:22).

of the “outer senses”, while time is the intuition of “inner sense.”⁸⁹ Space and time become fundamental structures, the necessary conditions, in terms of which we experience empirical subjects, but these structures are not of the world, but rather of human perspective – a subjective mode of representation of the mind or subject. Kant’s is a universal subjectivity, common to every mind. Phenomenal reality (spatial-temporal reality) is the only reality that is accessible to us.

We now have completely before us one part of the solution of the grand general problem of transcendental philosophy, namely the question – How are synthetical propositions *à priori* possible? That is to say, we have shown that we are in possession of pure *à priori* intuitions, namely, space and time, in which we find, when in a judgement *à priori* we pass out beyond the given conception, something which is not discoverable in that conception, but is certainly found *à priori* in the intuition which corresponds to the conception, and can be united synthetically with it. But the judgements which these pure intuitions enable us to make, never reach farther than to the objects of the senses, and are valid only for objects of possible experience (1781/1990:43).

Hence the possibility of the spatio-temporal framework of Newtonian physics could be certain only in terms of its being the structure of our experience of objects. Kant distinguished between *phenomena*, objects as we experience them; and *noumena*, objects in themselves, as they might be known by a pure intellect. Not having a pure intellect, our access to the world is restricted to knowledge of the phenomenal world, mediated by the senses. If we only have access to the phenomenal world, what would happen to our conception of the self or subject? One possibility that Kant discusses is a conception of the subject as phenomenon, experienced as an object conforming to the to the properties of space and time (39-41).

⁸⁹ This distinction is important in that it highlights a subtle difference in the scope of these two structures: time is the formal condition, *à priori*, of all phenomena, while space is limited as a condition *à priori* to external phenomena only (Kant 1781/1990:30).

Kant explains the possibility of synthetic *à priori* judgements through making this synthesis internal. He postulates internal categories of Understanding, through which we filter perceptions that we encounter in the external world. Human beings are given an active role in perception and the acquisition of knowledge. These categories of understanding become the transcendental conditions for knowledge - we cannot experience anything independently from them. For Kant sensations are already experienced as perceptions. To quote Levin (1992:36): "Kant moved the locus of knowing from the world to the mind."

Kant presents us with a mind that has three aspects: the Senses, Understanding and Reason. Kant subjects understanding to analyses in his transcendental logic. Where the senses contribute to mind intuitions of space and time, understanding contributes to mind general categories of conceptual schemata, according to which experience is then organised. Reason integrates knowledge and allows for self-awareness and self-criticism (Levin 1992:38). Kant proposes the existence of categories, which are the concepts that are necessary conditions for us to organise intuitions and make them accessible to judgement. He distinguishes between a number of categories of judgement, which he explicates in tables in the *Critique of Pure Reason* (1781/1990: 56; 62). The intellectual form of every judgement is characterised by quantity, quality, relation, and modality, each of which, in turn, contain three *momenta*. In all there are twelve categories or transcendental concepts of objects in general: three categories of quantity, namely unity, plurality, and totality; three categories of quality, namely, reality, negation, and limitation; three categories of relation, namely substance or inherence and subsistence, cause or causality and independence, and community or reciprocity between agent and patient; and finally three categories of modality, namely possibility, existence, and necessity. With these categories Kant tries to describe twelve ways of conceiving of an object that are necessary to make the twelve logical functions of judgement applicable to them. Through these categories it becomes possible to conceive of objects as substances, as standing in relations of cause and effect, and as parts of wholes in terms of which we can make judgements. These categories are *à priori* in origin and can thus be shown to be certain. The cost of this certainty is the realisation that

noumenally (independent from human thought) things represented to us might not obey the rules we attribute to them. On the positive side we now can be assured of the universal validity of the foundational principles of the Newtonian scientific world-view, in terms of the how things appear to *us*. Seeing that causality is a phenomenal property, attributed to experiences through Understanding, Kant concludes that noumena (including the self-in-itself), being outside the categories of space and time, are free from causality. By means of a transcendental deduction Kant aims to show that the application of his categories of thought to all possible experience is justified. As we will see, the existence of a continuous self becomes a precondition for Kant for much of our knowledge about the world.

As indicated earlier, Kant could not abide by Hume's conclusion that to render cognitions beyond the realm of experience he had to avail himself of the idea of habit. Kant speculates that Hume's error lay in his failure to consider that: "...the understanding itself might, perhaps by means of these conceptions, be the author of the experience in which its objects were presented to it (1781/1990:74). He argues that this empirical derivation cannot be reconciled to the fact (as established by him) that we do possess scientific *à priori* cognitions, namely pure mathematics and general physics. Kant endeavours to avoid the inevitable scepticism of Hume, by establishing the determinate limits, and the sphere of the legitimate activity of reason. Kant argues that that the senses cannot provide the conjunction of the manifold content of the representations they provide us with, and concludes that conjunction needs to be an act performed by our faculty of representation, in other words, the faculty of understanding.⁹⁰ In his own words: "Conjunction is the representation of the synthetical unity of the manifold" (76). Kant argues that the category of unity evidently presupposes conjunction, and he sets about looking "still higher" for the ground of this unity of diverse conceptions in judgment, therefore for the possibility of the existence of the understanding.

⁹⁰ Here, again, Kant stresses with regard to conjunction or syntheses, that nothing can be represented as a conjoined object that we (the subject) had not previously actively conjoined (1781/1990: 75).

Kant finds this unity in the *I think*, which must accompany all representations in the mind. He bases this assertion on the fact that if the *I think* did not accompany all representations, it would be possible for him to have something represented in him, without him thinking it, which he deems to be either impossible or at the very least, nothing in relation to him. Kant calls that the representation, the *I think*, pure apperception (his word for consciousness) and distinguishes it from primitive or empirical apperception.⁹¹ No representation can exist for me independently of the apperception. He calls the unity of the apperception the “transcendental unity of self-consciousness” (77). To quote Kant:

This relation, then does not exist because I accompany every representation with consciousness, but because I join one representation to another, and am conscious of the synthesis of them. Consequently, only because I connect a variety of given representations in one consciousness, is it possible that I can represent to myself the identity of consciousness in these representations; in other words, the analytical unity of apperception is possible only under the presupposition of a synthetical unity... [F]or the reason alone that I comprehend the variety of my representations in one consciousness, do I call them my representations, for otherwise I must have as many-coloured and various a self as are the representations of which I am conscious (78).

À priori synthetical unity is the foundation of the unity of consciousness, and antecedes determinate thought *à priori*. But this conjunction or synthesis does not belong to the objects of consciousness in themselves, but is the result of the operation of the understanding. Kant calls this the highest principle in all human cognition, where he is conscious of himself as a *necessary à priori* synthesis of his representation. The unity of consciousness constitutes the possibility of representations relating to an object, and the

⁹¹ The empirical consciousness that accompanies different representations is in itself fragmentary and has no relation to the identity of the subject (Kant 1781/1990: 77)

possibility of their objective reality; ultimately the unity of consciousness constitutes the possibility of the existence of the understanding itself, and therefore the objective condition of all cognition.

Guyer (1995:141-144) sums up this lengthy argument as follows: 1) all possible representations belong to a single, numerically identical self; (2) this is a synthetic connection of representations, which (3) requires an *à priori* synthesis among them, (4) the rules of which are non other than the categories, which are therefore (5) necessary conditions for the representation of any objects by means of the representations that themselves belong to a numerically identical self. Kant's "I" is the ground of the empirical unity of the self – not as a simple, spiritual being, but merely as a permanent substratum in time (Hatfield 1995:219).

We have seen that Kant sees the existence of the self as an enduring entity as a precondition for knowledge (Kant often equates self and consciousness, and self and reason) Hence Kant's answer to the question of the existence/ non-existence of the self is not, as with the empiricists, to look for the empirical confirmation of the self, through introspection for example. Empirically perceiving the self is impossible precisely because the self exists, as a precondition for coherent experience and perception. Levin states that the function of the self is synthetic, not in the Humean sense where habit and memory give us a sense of personal continuity and therefore identity, but as constitutive of experience – a sense of a unified self is logically prior to an experience as mine, and there can be no other experience (1992:38). By means of introspection it is also possible for one to experience the self as object, and as subject to time Kant (1781/1990:89).

Kant recognises the "paradox" that his view entails, namely that our intelligence provides us with a self as object, our self as phenomenon and not "as we are in ourselves" (88). Kant therefore postulates two selves, in that he distinguishes between a consciousness of self, and knowledge of self: the phenomenal self that is sometimes perceived in introspection and the noumenal self, which is inaccessible. The phenomenal self is knowable as the temporal sequence that is me. The noumenal self is the *I am* that transcendently accompanies every thought, but of which we cannot have any knowledge (90). Kant's noumenal self becomes the transcendental ego of

nineteenth and twentieth century philosophy, which is a purely logical condition of thought. By distinguishing between the noumenal and the phenomenal self, Kant preserves the possibility of free will. He does believe in two realms, a realm of necessity (the phenomenal realm) and a realm of freedom (the noumenal realm). The noumenal realm is not subject to the causality, which characterises the phenomenal realm, and it follows that the noumenal self is free from causality. This enables man to stand outside the causal chain events and exercise free will. The noumenal self is seen as outside of the realm of logical possibility and thus free and even potentially immortal. Kant manages to reintroduce a conception of soul, but one that is not derived from, what Levin calls, "the illegitimate use of reason" as has been the case in metaphysics up to that point (Levin 1992:40).

Kant managed to change the project of modern philosophy. His recognition that there was a connection between knowledge of the self and knowledge of objects, he undermined the Cartesian idea that we could have knowledge of our inner sense without any knowledge of the empirical world. He also undermines Hume's project of grounding all knowledge in empirical experience. Kant established knowledge as the result of judgement, which necessarily involves empirical input *and* internal logical structures.

5. Fast forward to the twentieth century

The legacy of the Enlightenment would prove to be the entrenchment of the mind-body split in Western philosophy. Over the subsequent three or so centuries the mind would stay inextricably intertwined with reason, however limited or qualified a given conception of reason might be. The self would continue as something to be examined and understood through introspection and self-examination. The conception of the rational mind as logical precondition to thought would be modified in that time, but would essentially go unchallenged until the end of the nineteenth centuries.

The turn of the twentieth century heralded a famous turning-point in the conception of mind that would have dramatic implications for the reigning conception of mind-body at all levels. This transformation carried the name of Sigmund Freud. Freud played a pivotal role in dispersing the idea that the mind and the body are two distinct, independent entities, as well as calling

into question the idea of the rational mind that had held sway for so many years, by postulating the idea of the *unconscious*. The implications of this important development are so far-reaching that Freud merits being discussed at length. The following chapter will aim to do just that, as well as to link Freud's conception of the way that the mind operates with contemporary models for the functioning of the brain.

Chapter Three

A Neurological Basis for the "Higher Functions" of the Brain

It is easy to speculate, but hard to confirm, that Freud's expansion of the bounds of the thinkable was a pre-condition for a much more pervasive, and much less controversial, style of theorizing in experimental, and especially cognitive, psychology in recent years.

Dennett (1998: 162)

The new mentalism, combining tenets from previously conflicting views, tends to reconcile polar opposites of the past such as mind and matter, the physical and metaphysical, determinism and free choice, as well as "is" and "ought" and fact and value, in a unifying view of mind, brain, and man in nature.

Sperry (1998:165)

It is only by means of such complicated and far from perspicuous hypotheses that I have hitherto succeeded in introducing the phenomena of consciousness into the structure of quantitative psychology.

Freud (1950: 311).

Whereas two or three centuries ago it seemed obvious that the sensations and images, thoughts and voluntary impulses, were kinds of "capacities", or immediate functions of specific brain organs - thinking of the brain as a system of specialized "micro-organs"- such a concept is no longer acceptable. It is better to suppose that mental processes are complex information-processing activities, reflecting reality. Instead mind is now considered to be a product of active processing of the flow of information working through elementary drives, or complex motives, set to single out important information about reality, relating bits of information and synthesizing them, and establishing plans and programmes of behaviour, and in conscious control of actions.

Luria (1998:489)

1. Introduction

By proposing the existence of the unconscious Sigmund Freud turned many received wisdoms about the mind on their head. The possibility arose that the "higher functions" of the brain need not be conscious processes, and that there was much more to the "mind" than met the introspective eye. Far from being, clear, "certain, and indubitable" (to use Descartes' phraseology), conscious (rational) thought processes became entangled in the intricacies of

the physiological functioning of the body. Freud introduced a new way of thinking about man and about mental functioning (Padel 1998:270). Many of his concepts and ideas are commonplace today, but must have been disconcerting to his contemporaries (Zangwill 1998a:268-270), think of examples such as *repression*, *transference*, *fantasy*, *fixation*, *Oedipus-complex*, *free association*. His theories had an enormous influence on conceptions of personality, and its development. Much emphasis was placed on understanding the present in terms of the past; experiences in infancy would lay the foundation for the adult personality. Freud's theories allowed for conflict within the psyche and incorporated a host of unconscious motivating factors into the thought process. Influences on thought had everything to do with primal drives, and very little to do with rational decision-making. For him all mental life is originally unconscious, and only has the potential to become conscious when one adapts to external reality, i.e. when one develops one's abilities of perception, learning and language. Freud was convinced that eventually psychology would be based on organic, or bodily principles (Zangwill 1998b:277).

Freud's ideas were groundbreaking, and as such encountered their fair share of resistance. Although he seemed to lay the foundation of a theory of mental processes that would overcome the difficulties posed by dualism, Freud never explicitly addressed the issue. In fact, the conception of mind and body as consisting of different "substances" and subject to different causal laws would prove to be tenacious, and many current theorists still adhere to dualist principles, although they would probably not call themselves "dualists". In some circles, denying self/mind as a mystical, non-material entity is received as an abrupt dismissal of the very existence of self/mind. Kenny (1989) for example, dismisses the concept of "self" as nothing more than a grammatical error. Kenny's position will be briefly discussed in the following section. It will prove useful in serving as a counterpoint to Freud's ideas. Freud's *Project for a Scientific Psychology* is of particular interest to this study, in that it highlights some of the presuppositions inherent in dualist theories of mind, and it provides an alternative approach to mind, which can overcome some of the difficulties posed by dualism. The Project and its

influence on later Freudian theories will be discussed in Sections 3 through 7. First, however, we will give a brief overview of Kenny's dismissal of the concept of self as the result of linguistic misunderstanding.

2. The self as a grammatical error

Kenny believes the self to be a "mythical entity" (1989:87). Describing it as a piece of philosopher's nonsense, a misunderstanding of the reflexive pronoun, nothing more than a grammatical error:

To ask what kind of substance my *self* is like asking what the characteristic of my *ownness* is which my own property has in addition to being mine. When, outside philosophy, I talk about myself, I am simply talking about the human being, Anthony Kenny, and my self is nothing other than myself. It is a philosophical muddle to allow the space which differentiates between "my self" and "myself" to generate the illusion of a mysterious metaphysical entity, distinct from, but obscurely linked to, the human being who is talking to you (87).

Kenny believes this grammatical error to be the erroneous belief that 'I' is a referring expression (88), but he insists that the erroneous belief in the self is not the result of grammatical error alone, but that it has two other roots: one epistemological and one psychological. Predictably, the epistemological error has its roots in Cartesian scepticism, for much the same reasons as given by Ryle. How can the non-material ego act upon the material realm of the body? What relation can be shown to exist between the mental substance and its transient conscious thoughts? Kenny argues that what Descartes describes as indubitable thought can at most be applied to one's own imagination, and not with one's "intellectual" mind (90).

The psychological root of the deluded conception of a self lies with a misconception in the empiricist tradition, most notably with Locke, where again, according to Kenny, an illegitimate distinction is made between the intellect and the imagination. The misconception here is the idea of the self as

the subject of inner sensation (although Kenny does allow that Hume concluded that the self was an illusion). Kenny argues that the self of modern philosophy is a chimera, stemming from empiricist error. From here the prevailing modern notion of the self as essentially a perspectiveless subject of experiences where the self cannot easily be identified with particular bodies (92).

Kenny's objection to the idea of a self becomes more understandable when one gains clarity on what he believes the concept to denote: a self is something different from oneself, it is something over and above human being, something other than body and mind (93). He argues that it is possible to think objectively about oneself, just as it is possible to make "impersonal, centreless" scientific judgements (94) and argues that the possibility of making objective scientific judgements about oneself negates the necessity for postulating the concept of "self." How objective judgements about self are to come about is not made clear.

Kenny is not at all convincing in his attempt to prove the concept of self a philosopher's myth. What is striking in his essay is, once again, the intellectual quagmire one is lead into when discussing a concept (entity?) as ambiguous as the self. Kenny's blithe distinction between the mind and body, and the self as something over and above those entities, something quite mystical, speaks volumes about many different conceptions of self. Far from undermining the rationalist and empiricist traditions, Kenny only seems to be yet another player in a long-standing family dispute. One need not be caught up in an either/or situation. Denying the mysticism of the self need not necessarily imply that the self, in some or other form, does not exist at all.

Although he argues that objective judgements about oneself are possible, Kenny is very sceptical about the possibility of orchestrating a scientific study of the mind. Mainly because, he argues, we cannot identify states of mind with physical states of the brain. This is a common objection against the possibility of a "material" mind, among philosophers of mind.⁹² Kenny's main objection seems to be the extension of the determinism of

⁹² See our discussion on Nagel's "What is it like to be a bat?" pp. 177-181.

science to the operation of the mind (140). A science of mind would then presumably have the capacity of accounting for both the operation and the origin of the mind in terms of the principle of determinism.⁹³ Kenny differentiates between psychological and non-psychological determinism, and argues that psychological determinism incorporates "mentalist" terms, i.e. terms for "mental events" and "states of mind". He argues that trying to account for someone's actions in a psychological deterministic way is illegitimate, because this approach erroneously equates reasons with causes. In other words, the psychological determinist treats mental states, such as wants and beliefs, as being causally connected to physiological processes (143). And Kenny concludes that: "All psychological forms of determinism are incoherent because they misconstrue that nature of the mental phenomena to which they explicitly or tacitly appeal in their formulation" (145).

Kenny does consider the possibility of determinism as it pertains to the neurophysiological states of the brain and the central nervous system. Although he cannot find fault with this theory on grounds of internal incoherence as with psychological determinism, Kenny rejects it on account of its implications for our conceptions of human freedom.⁹⁴ (Indeterminism or randomness could not possibly amount to free will!) Contrary to what one might expect, Kenny does allow that human freedom is possible, in spite of physiological determinism (148). The reason he allows for this possibility is that he does not see a reason to believe that physiological determinism entails that a particular physiological event needs to be correlated with particular psychological conditions in a regular and law-like manner. Kenny argues that it is possible that a particular want may at different times be correlated with *different* physiological processes. He concludes that

⁹³ Kenny limits his understanding of determinism to the following general scheme: "...if determinism is true, it will be the case for any event E that there was an antecedent event or state C such as there is a covering law to the effect that whenever a situation such as C obtains there will follow an event such as E. Every event will fall under a description such that there exists a law from which, in conjunction with a description of the antecedent conditions, it can be deduced that an event of that description will occur" (1989:141).

⁹⁴ Kenny (1989: 148) defines freedom as having the power to do otherwise: Freedom undoubtedly involves the power to do otherwise. I do X freely only if I have the power not to do X, and that means I have the opportunity not to do X, and the ability not to do X.

conceding the existence of physiological determinism need not imply predictability at a psychological level, which leaves the possibility of human freedom intact.

Despite this conclusion Kenny declares himself to be agnostic on the issue of determinism :

The issue of compatibilism [of determinism and free will] is a strictly philosophical issue: it is a question about the logical relationship between two sets of concepts. But on the assumption that determinism can be coherently formulated, the issue between the determinists and the indeterminists is not a purely philosophical question. The question concerns the nature of the system of laws governing the universe. If this question can be answered it cannot be answered by the philosopher alone. It is an issue on which the philosopher as such can and should remain agnostic (150).

Ultimately, however, Kenny does reject the materialist idea that mental states and structures are simply physical states and structures "described at a certain level of abstraction" (151). He does not believe there to be a one-on-one correlation between physical and psychological, or mental and physical, states. He insists that "the physical object which is described by *mentalistic* predicates is a human being, not a human brain."

If my brain were as deterministic as an electronic computer, so that its entire output could be predicted from the inputs it receives, that would not suffice for anyone be able to predict the thoughts that I will have. For what gives meaning to any kind of output of my brain – whether channelled through action, speech, or writing – is something which is quite external to it, just as what gives meaning to the output of the computer is external to it (153).

He also suggests that to see the mind as a purely physiological entity implies that "contents" of the mind, like language, must have evolved through a process of natural selection.⁹⁵ In the end it does not strike one that Kenny finds a science of mind a useful or even desirable enterprise. The main reason seems to be the importance that he attaches to upholding the distinction between the "mentalistic" and the physical. Undermine the distinction, and the possibility of a science of mind seems all the more plausible. If one does not see the need to ascribe a different status to states of mind from that of possible neurological underpinnings, perhaps one can postulate a coherent, materialistic theory of mind. Such a materialism need not necessarily lead to an extreme form of determinism, or fatalism and leave human beings without freedom of will. Equating the human being with the human brain need not in any way detract anything from being human. A materialist account of the mind might serve to demystify "mental" phenomena, but such an event could only serve to enhance our understanding and appreciation of the intricacies of the human mind - and that is precisely what Freud sets out to do in his *Project for a Scientific Psychology*.

3. Freud's "new mind"

The *Project for a Scientific Psychology* (1950 [1895]) was one of Freud's earliest works in which he, essentially, creates a model for the neurological functioning of the brain. His aim was to establish psychology as a natural science by representing "psychical processes as quantitatively determinate states of specifiable material particles" (1950 [1895]: 295). Although this project was abandoned in frustration, hampered by a lack of neurological information (Freud 1915c: 174-176) (neurology was then in its infancy), the *Project's* influence would be felt throughout Freud's

⁹⁵ Although I do not discuss Kenny's view on language in detail, the kind of arguments that he uses and assumptions that he makes, makes for very interesting philosophical debate. See for example Cilliers (1989: 152-199) and (1998:37-47 and 123-126) where he highlights some of the possible interactions between language and consciousness and discusses post-structural language theories, which serves to undermine many of Kenny's presuppositions.

psychological works (Strachey 1986:290).⁹⁶ What concerns us here is not so much the debate on whether or not Freud's speculations can indeed be regarded as precursors to contemporary neurological theories, but rather the revolution in thinking on psychological (mental) matters that Freud's work personifies and doubtlessly instigated. Here we have a systematic model that explains the "higher mental functions", like consciousness, as a result of neuronal functioning. Even though Freud abandoned his attempt at modelling the neuro-physiological characteristics of the mind, he never completely abandoned the belief that psychology would one day be explained in terms of physiological functions of the brain (Zangwill 1998b: 277). Freud's analyses of the conscious and the unconscious would also presage a radical departure from conceptions of that most important of human traits as it has been discussed up to this point: consciousness.⁹⁷

In the rationalist/empiricist debate, nothing is more central to the mind than consciousness. All the activities of mind are accessible to itself, and by means of introspection the mind is able to observe and speculate upon its own attributes. This view was accepted as self-evident, a *prima facie* necessity for the concept of mind. So much so that Sigmund Freud's hypothesis that something like unconscious mental processes might exist was rejected as a conceptual impossibility. Dennett (1998:162) suggests that initially Freud was able to win converts to his theory, by allowing for the possibility that unconscious mental processes could be described as belonging to other "selves" within the psyche. In other words, the possibility of splitting the subject into many subjects, one preserves the possibility that every mental state is "someone's" conscious mental act. While Freud could

⁹⁶ See the editor's introduction to the *Project for a Scientific Psychology* (1950 [1895]) in the Standard Edition of the Complete Psychological Works of Sigmund Freud, pp. 283-293 for a brief discussion on both the similarities and differences between the Freudian and more contemporary approaches to physiological explanation of mental processes. The editor also voices his concern that latter-day theories may be over-hastily attributed to Freud's somewhat obscure formulations.

⁹⁷ Sternberg specifically links Freud's "revolutionary" ideas to the prominence of thermodynamics in the physical sciences. Sternberg, like Prigogine, draws parallels between the processmatic nature of thermodynamics and the emphasis placed by Freud on the dynamic processes underlining the personality. This new psychodynamic approach would place great emphasis on biological drives and processes, which rose to prominence thanks to the work done by Darwin (1995: 597-598).

claim that, given his clinical observations, he was able to override patient's assertions about what was happening in their minds, he could also conclude that sophisticated processes of reasoning were happening that were entirely inaccessible to introspection on the part of the subject.

Dennett goes on to argue that the growing acceptance of the idea that unconscious processes were in fact the non-conscious "information processing" of organic machinery paved the way for the other extreme: calling the very necessity and existence of consciousness into question.

Freud's concept of the unconscious as part of the structure of the mind that operates outside of the awareness of the subject and can radically influence conscious processes would completely alter our conception of the mind. Since then, no examination of the mind could be complete without taking the unconscious into account. The possibility of phenomena like repression and resistance stemming from the unconscious, and having a compelling influence no matter the will of the conscious subject, paved the way for a dynamical theory of the mind, where mental forces can be in conflict with one another (See Strachey 1986:19 and Sternberg 1995: 598). Freud would famously obtain his subject matter by means of, among other things, self-analysis and also through the interpretation of dreams.

4. The Project

Freud's intention with this work, as his title suggests, is to establish psychology as a natural science. He aims to do this through representing psychical processes as "quantitatively determinate states of specifiable material particles, thus making those processes perspicuous and free from contradiction" (1950 [1895]:295). Freud's point of departure for his neurological model is based on two principles: i) the notion of quantity (Q)⁹⁸ and ii) the nervous system consists of interconnected material particles called

⁹⁸ Quantity is defined as "what distinguishes activity from the rest" (read: energy or neuronal excitation) and is described as being *subject to the general laws of motion* (Freud 1950 [1895]: 296).

“neurones.” Freud envisions the nervous system to have the following structure:

[T]he nervous system consists of distinct and similarly constructed neurones, which have contact with one another through the medium a of foreign substance, which terminate upon one another as they do upon portions of foreign substance [and] in which certain lines of conduction are laid down in so far they [the neurones (sic.)] receive [excitations] through cell-processes [dendrites] and [give them off] through an axis-cylinder [axon](298).⁹⁹

Freud conceives of “neuronal excitation” as quantity¹⁰⁰ (Q) in a state of flow, and consequently the principle postulates the principle of neuronal inertia: neurones tend to divest themselves of Q. This principle is then used to explain the structure, functions and development of neurones (296). Neurones are connected with one another through permeable contact-barriers, which allows Quantity the ability to flow between neurones. Freud believes that nature tends towards equilibrium; he calls this the *principle of constancy*. A neurone “filled” with Q is in a state of tension and would, in accordance with this natural tendency, “tend to divest [itself] of Q”(296).

Quantity (Q) is acquired through the sensory apparatus and discharged by the neurones to the muscular mechanisms, in an attempt to keep themselves free of the external stimulus and regain equilibrium, and hence

⁹⁹ Compare this account to a contemporary description of the basic structure of the nervous system:

The nervous system is made of cells, like every animal tissue. The essential cell is a nerve cell or neurone. The neurone is considered to have three parts: the cell body, the dendrites and the axon. The dendrites are thin prolongations of the cell body. Most sorts have one prolongation far longer than the others; this is the axon, the telegraph wire of the neurone, taking the message from one neurone to another or else to the muscle or gland it supplies. The boundary of the neurone is the membrane, having certain properties on which the functioning of the nervous system depends... (Nathan 1998:514).

¹⁰⁰ Freud distinguishes between two “types” of Q:
 Q = Quantity (in general, or in the order of magnitude in the external world), hence where energy is received from the external world through the senses.
 Q_n = Quantity (of the intercellular order of magnitude); a kind of internal energy.

not upset the principle of inertia (296). Crudely simplified, it is this quantity that stimulates the nervous system, and the divestment of quantity through the nervous system to the muscular system, which enables one to react to a state of affairs in the external world. With this model Freud is able to explain stimulus/ response and reflexive behaviour. Discharge of Q is seen as the primary function of the nervous system.

There is, however, another source of energy, which also upsets the principle of inertia and is not so easy to divest. $Q\acute{n}$ can be explained as a form of endogenous energy.¹⁰¹ This energy has its origin in the cells of body itself, especially in the basic needs of the body – hunger, respiration and sexuality (297).¹⁰² Although $Q\acute{n}$ also obeys the principle of tending towards equilibrium, it cannot be discharged without the body's needs being satisfied. If external conditions are not conducive to meeting bodily needs, this energy cannot be discharged and builds up in the nervous system. In unfavourable conditions, the nervous system cannot meet the constancy principle by discharging $Q\acute{n}$. The nervous system is obliged to abandon its trend toward inertia, and must accommodate this store of energy. A *cathected* neurone is one filled with $Q\acute{n}$. The trend does persist, however, in that it is "modified to keep Q as low as possible and to guard against any increase of it," in other words, to keep energy in the system constant (279). Accumulating and keeping constant $Q\acute{n}$ is seen as the secondary function of the nervous system.¹⁰³

The key to explaining how this flow and regulation of Q is possible lies in the permeability of the contact barriers that exist between neurones. It must be remembered that any one neurone is connected to many other neurones and there are therefore many possible routes that energy flowing through the nervous system could take. Energy is more likely to pass through barriers that

¹⁰¹ See footnote 100.

¹⁰² See Freud's *Instincts and their Vicissitudes* (1915a: 117-140) where he elaborates on this basic scheme and establishes instinctual stimuli as stimuli arising from the organism itself, operating as constant forces.

¹⁰³ In Section vii of the *Interpretation of Dreams* (1901) Freud elaborates on the primary and secondary processes, especially linking their operation to that of the unpleasure principle (later renamed the pleasure principle), wishing and repression (see 598-601).

have lower resistance. Freud makes the fundamental observation that the more energy passes through a barrier, the lower its resistance will become, allowing *pathways* to form through the system of neurones. This forms the basis for the formation of memory: such pathways, once formed, will tend to remain constant in the system.¹⁰⁴

Memory, in its turn, becomes the basic property of neurones and their interconnections.¹⁰⁵ With the fixing of pathways in the neuronal system as a result of the passage of quantity, Freud needs to be able to account for new information being added to the system, without the system becoming saturated. He does this by distinguishing between two types of neurones: perceptual cells (ϕ neurones), and mnemic cells (ψ neurones). Perceptual cells are permeable, in contact with the outside world and can transport quantity without changing state, while mnemic cells are impermeable, in contact with the body and only allow the passage of quantity with difficulty. The latter neurones have the capacity of representing memory. Memory is seen to be the result of “the facilitations existing between ψ neurones”(300). Cilliers (1989: 112) emphasises the fact that Freud makes an important realisation at this point, that is, if all contact barriers between all neurones were to be equally well facilitated the neuronal structure would be completely homogenous, making memory impossible. There would be no reason why

¹⁰⁴ With this theory Freud has in actual fact proposed a way in which a mass of undifferentiated neurones can acquire structure through *self-organisation*. The reader will recall that the concept was discussed in Chapter 3, as a characteristic of an open, complex system. Through self-organisation becomes possible to give a plausible explanation as to how the mental apparatus can develop structure, without being pre-programmed, and by taking environmental influences into account. Freud's basic explanation of how neuronal pathways are established corresponds with contemporary theories, particularly Hebb's famous *use-principle*. According to this principle the connection strength of a synapse between two neurones should increase proportionally to how often it is used (Cilliers 1998:17). The stronger pathways' synapses become more effective and hence these pathways are used more often, while unused pathways wither away (Young 1998:455). In this way a mass of largely undifferentiated neurones can develop a structure that is based on the information available to each neurone locally. In other words, the networks of neurones can learn through experience.

These characteristics are essential to revised conception of the self as developing and self-organising within a system of differences and will be returned to again when we discuss both contemporary neurology and the self as an open system.

¹⁰⁵ Freud defines memory as: “a capacity for being permanently altered by single occurrences” (1950[1895]: 299).

one neuronal pathway would be preferable to another. Freud takes the crucial step of redefining memory as “represented by the differences in the facilitations of ψ neurones” (1950 [1895]:300). As Cilliers puts it: “Memory does not lie in the facilitated pathways themselves, but in the relationship between them, and this relationship is one of differences” (1989:112).

Freud postulates memory as the basic component of the nervous system and as prior to consciousness and cognition. The procedure of forming memory traces is entirely unconscious, and with the following statement Freud launches us into the realm of unconscious mental processes:

Hitherto, nothing whatever has been said of the fact that every psychological theory, apart from what it achieves from the point of view of natural science, must fulfil yet another major requirement. It should explain to us what we are aware of, in the most puzzling fashion, through our “consciousness” and, since this consciousness knows nothing of what we have so far been assuming – quantities and neurones – it should explain this lack of knowledge to us as well.

We at once become clear about a postulate which has been guiding us up to now. We have been treating psychical processes as something that could dispense with this awareness through consciousness, as something that exists independently of such awareness. We are prepared to find that some of our assumptions are not confirmed through consciousness. If we do not let ourselves be confused on that account, it follows, from the postulate of consciousness providing neither complete nor trustworthy knowledge of the neuronal processes, that these are in the first instance to be regarded to their whole extent as unconscious and are to be inferred like other natural things (Freud 1950 (1895): 307-308).

5. Freud's differentiated neurons become conscious

Freud needs to make room for consciousness in his neuronal system. He goes about this by first by claiming that we are not able to be conscious of quantity (Q), but only of quality – sensations which are different in many ways, and whose differences are distinguished according to their relation with the external world, but are wholly independent of quantities (308). Qualities do not originate in the external world, where there are "only masses in motion and nothing else"(308). Nor do qualities originate in the ϕ , and ψ systems, which perform processes that are without quality. Freud speculates that there must be a third system of neurones – ω neurones – whose states of excitation give rise to various qualities, in other words, conscious sensations (309).

If we keep firmly to the fact that our consciousness furnishes only *qualities*, whereas science recognizes only *quantities*, a characteristic of the ω neurones emerges, as though by rule of three. For whereas science has set about the task of tracing all the qualities of our sensations back to *external quantities*, it is to be expected from the structure of the nervous system that it consists of contrivances for transforming external *quantity* into quality; and here the original trend to keep off *quantity* seems to triumph once more (309).

Neither the primary, nor the secondary processes are under the control of a conscious ego. Freud concludes that quality (conscious sensation) comes about when quantity is as far as possible excluded. Although still cathected with Q_n and still striving towards discharge, the ω neurones are moved by minuscule quantities. However, Freud does not conclude that the ω neurones are therefore even more impermeable than the ψ neurones (given that permeability depends on the effect of Q_n). These "vehicles of consciousness" (ω neurones) need to be completely permeable, and they have to be able to return to their former state. In other words, these neurones require complete facilitation and permeability if they are to accommodate the characteristics of consciousness. Given that quantity is virtually absent in the ω system, this

permeability and facilitation must arise in a way other than the possessing of Q .

Freud finds his way out of this difficulty by ascribing a temporal characteristic to the flow of neuronal energy (over and above its spatial arrangement) – temporality¹⁰⁶ (310). He refers to temporality as *period*. The resistance of contact barriers now does not depend only on the transference of Q , but also on the period of this neuronal motion, which is distributed in all directions, without any hindrance. The fact that the ω neurones are affected by period, rather than by Q is then taken to be the fundamental basis of consciousness.

As Cilliers (1989:116) points out, with this proposal Freud merely manages to shift the question of what consciousness is to another level. Consciousness is now a function of a specific system, instead of being integrated with the rest of the nervous system. Freud himself seems uncomfortable with this idea and proposes an alteration in a letter to his long-time correspondent Fliess in 1896:

I now [in my new scheme] insert these ω neurones between the ϕ neurones and the ψ neurones, so that ϕ transfers its quality to ω , and ω now transfers neither quality nor quantity to ψ , but merely excites ψ - that is, indicates the pathways to be taken by the free ψ energy (388).

Cilliers sees this move as vitally important in that it is the first step towards dropping the ω system, and seeing consciousness not as the property of a specific system, but as the result of *interaction* (i.e. the interaction between perception and memory) (116).¹⁰⁷

¹⁰⁶ Freud substantiates his inference on the grounds that the "mechanics of the physicists have allowed this temporal characteristic to the other motions of mass in the external world as well" (310).

¹⁰⁷ Cilliers (1989:116) contends that in *Note on a "Mystic Writing Pad"*, Freud seems to formulate a theory of consciousness, without recourse to ω neurones, as we will discuss.

Freud takes note of two other theories of consciousness: 1) the mechanistic theory, which sees consciousness as a mere epi-phenomenon of "physiologico-psychical processes" and 2) a subjectivism of sorts, where consciousness is the subjective side of psychical events, and thus completely reliant on specific physiological mental processes (311). Freud's theory falls somewhere in the middle, where consciousness is partly subjective and partly the result of external conditions (Freud 1950 [1895]: 311; Cilliers 1989: 117). If consciousness is omitted, psychical events would be altered, specifically the contribution from ω .¹⁰⁸

Having established the basic mechanism for generating consciousness, Freud proceeds to explain a number of the "higher" psychological functions. It will not serve the purpose of this chapter to examine them all in detail, but the ones significant to our discussion will briefly be examined.

Freud postulates that the accumulation of $Q\acute{n}$ will create an urgency in ψ to discharge this energy. The only way for the pressure to be relieved is to affect an external change to get rid of the stimulus that causes the release of $Q\acute{n}$ in the interior of the body. Initially (in infancy) the human organism is incapable of bringing about the necessary action, and it has to take recourse to external help, in the form of a caretaker. The child's internal state needs to be communicated to the caretaker, which is accomplished "by discharge along the path of internal change" (1950 [1895]:318). Freud believes that this initial helplessness of human beings is the *primal* source of our moral

¹⁰⁸ As far as the content of consciousness goes, Freud speculates that, besides the sensory qualities that he has already attributed to consciousness, it also exhibits a series of very different sensations: those of *pleasure* and *unpleasure*. Freud links the sensations of pleasure and unpleasure to the increase (unpleasure) and discharge (pleasure) of $Q\acute{n}$ in ω , as influenced by the processes in ψ :

The aptitude for perceiving sensory qualities which lie, so to say, in the zone of indifference between pleasure and unpleasure disappears with the [presence of the] feeling of pleasure and unpleasure. This might be translated: the ω neurones show an optimum for the receiving the *period* of neuronal motion at a particular [strength of] cathexis; when the cathexis is stronger, they produce unpleasure, when it is weaker, pleasure – till, with a lack of cathexis, their capacity for reception vanishes (Freud 1950 [1895]:312).

motives.¹⁰⁹ He does not elaborate on the relationship between the infant and the caretaker, nor does he elaborate how this affects our subsequent moral motives, but the meaning he attaches to this initial and inevitable encounter seems to be that social interaction is a necessary precondition for the development of consciousness.

Having had his needs met by the helpful external agent, the infant is able to remove the endogenous stimuli and so experience satisfaction. Freud declares that this process has a radical influence on the development of the individual's functions.¹¹⁰ The motor image generated in ψ by sensory excitation allows for reproductive remembering. Contact barriers are facilitated between ψ neurones when energy is discharged due to needs being met. *Qñ* passes more easily between facilitated neurones than between ones that are not facilitated. These contact barriers ensure that consciousness would move from a stimulated neurone to one remembered to previously have been at the same time stimulated. When a certain urgency is experienced once again, the relevant memories are triggered.¹¹¹ As Cilliers (1989:118) points out a second conclusion that can be drawn from Freud's proposed process is that the body (and its needs) is a necessary precursor to consciousness. Bodily needs, or endogenous stimuli, become one of the primary movers of the physical system.¹¹²

The residues left by the experience of satisfaction or pain are what Freud calls affects or wishful states, which, Freud suggests, leave behind motives which are of a "compulsive kind" (1950[1895]:323). Wishful states are encountered when a positive attraction is felt toward the mnemonic image of a wished-for object (primary *wishful attraction*), while an object which caused or

¹⁰⁹ This should serve to refute the allegation of solipsism that is sometimes levelled at Freudian theory.

¹¹⁰ See also section vii of *The Interpretation of Dreams* (1901).

¹¹¹ Freud returns to this topic again in *The Interpretation of Dreams* (1901), *Beyond the Pleasure Principle* (1920), and *A Note on the "Mystic Writing-Pad"* (1925).

¹¹² In *The Interpretation of Dreams* Freud remarks on research done by Strumpell, from which comes the conclusion that one falls asleep when all one's most important sensory channels are closed. However, we cannot ward off the onslaught of stimuli completely and the stimuli that are not of sufficient strength to wake us, are then likely to be instigators of dreams (1901:24).

heralded pain will cause revulsion (primary *defence*) and the mnemonic image is abandoned as soon as possible (322).¹¹³

Freud then proceeds to link the mechanisms of both wishful attraction, and repression to the existence of the ego. He argues that the existence of both of these processes indicates that there exists an *organisation* in ψ , which can interfere with the passages of $Q\acute{n}$. This organisation is called the *ego*. His formal definition reads as follows:

Thus the ego is to be defined as the totality of the ψ cathexes, at a given time, in which a permanent component is distinguished from a changing one. It is easy to see that the facilitations between ψ neurones are a part of the ego's possessions, as representing possibilities, if the ego is altered, for determining its extent in the next few moments (1950[1895]:323).

The ego is responsible for the secondary processes in the mental apparatus. The "endeavour" of the ego is to "give off its cathexes by method of satisfaction" through means of inhibition, or influencing the repetition of experiences.¹¹⁴ Where side-cathexis has taken place, in other words if a neurone adjoining a cathected neurone is simultaneously cathected it acts to modify the course of $Q\acute{n}$, which would otherwise have been directed towards an already facilitated contact-barrier. The ego, the existing cathexes in ψ at a given time, acts to inhibit the passage of $Q\acute{n}$ from a hostile mnemonic image and so suppress unpleasure. ψ cannot distinguish between a wishful state and reality and a further function of the ego is to distinguish between a perception and memory. In his discussion on the properties of dreams Freud is careful to

¹¹³ Refer to *Project for a Scientific Psychology* (1950 [1895]:322) and *The Interpretation of Dreams* (1901:546) for a more comprehensive discussion on the mechanisms that drive both *wishful attraction* and *repression*.

¹¹⁴ Later on Freud would speculate that sleep can be explained by means of these mechanisms (1950[1895]:336). After the satisfaction of demands bombarding the nervous system from endogenous stimuli, the ego is unloaded and becomes temporarily superfluous and the individual enters a state of inertia - sleep. Freud describes the will as "the discharge of the total ψ $Q\acute{n}$ " (337). Sleep is characterised by an absence of will.

emphasise that consciousness does not cling to the ego, but can be associated with “any ψ process” (340).

In his *Formulation on the Two Principles of Mental Functioning* (1911) Freud again takes up these themes, applying them to his clinical findings. His discussion is not cast in neurological terms, however, and here primary and secondary processes are discussed on a purely “psychological” level. Nevertheless the influence of his fundamental neurological hypotheses is unmistakable and his discussion here, as well as in *The Interpretation of Dreams* (1901), serves to augment and develop many of the themes conceived in the *Project*. He observes that neurosis alienates a patient from a reality which the patient finds unbearable (1911:218). This alienation or turning away from reality is accomplished by means of repression.

Freud seeks the origin of neuroses in the “unconscious mental processes”— the primary processes which he describes here as “the residue of a phase of development in which they were the only kind of mental processes” (219). The governing principle behind these processes is the pleasure-unpleasure principle (later abbreviated to the pleasure principle). The psychical activity that draws us back from experiences that might cause us unpleasure is called repression and the remnants of this principle explains both dreams and our waking tendency to distance ourselves from disagreeable impressions.¹¹⁵ In both the *Project* and *On the Interpretation of Dreams*, Freud had developed his ideas on the wishful state, where endogenous demands are originally met in a hallucinatory manner. On the event of the disappointment that is encountered when internal needs are not met by means of these wishful hallucinations, the psychical apparatus develops a conception of what is really the case in the external world, rather

¹¹⁵ In *Repression* (1915b:147) Freud explains the essence of repression as: *turning something away, and keeping it at a distance from, the consciousness*. What is primarily repressed is not withheld from consciousness altogether, but makes its way into consciousness if it has been sufficiently distorted and distanced from its original representation.

than forming a presentation that is merely agreeable. This mental function which is introduced is the *reality principle*.¹¹⁶

In the reality-detection game *consciousness* becomes a significant role-player. The consciousness needs to develop a sense for sensory qualities along with those of pleasure and unpleasure. Consciousness is augmented with the function of *attention* for precisely this purpose. Attention, on its part, depends on the existence of a faculty of *memory*. Consciousness is then in a position to make impartial judgements on the truth or falsity of given ideas, by comparing the idea to memory-traces of previously encountered reality and deciding whether the given idea is in agreement with reality or not. Motor discharge is now allotted with the function of action, in other words with the task of altering reality.¹¹⁷ As we have already seen in our discussion of the *Project*, *thinking* developed from the presentation of ideas, and arose out of the need to allow the mental apparatus to tolerate the increased strain put on it, while discharge is postponed.¹¹⁸

6. In the realm of the unconscious

In the *Project* (1950 [1895]:374) *Two Principles of Mental Functioning* (1911:221), *The Unconscious* (1915c:196-204) and *The Interpretation of Dreams* (1901:7-21;48-65) Freud emphasises his conjecture that thinking was originally an unconscious process and that it did not acquire qualities other than the ability to compare the relation between impressions of objects until it became connected with "verbal residues".

¹¹⁶ With the introduction of the reality principle one source of pleasure in our thought-activity remained that still obeyed the pleasure principle and did not succumb to the reality principle: fantasising (or *phantasying*, as Freud has it) and *day-dreaming*. The sexual instincts also remain detached from this development and develops a unique relation to the pleasure principle (cf. *Formulations on the Two Principles of Mental Functioning* 1911:222).

¹¹⁷ Initially Freud attributed to motor discharge the function of relieving the neuronal structure of accumulated stimuli through causing movement in the body to express the "unpleasure" created (1911: 221).

¹¹⁸ Hallucination as a means of dealing with accumulating stimuli is no longer a viable option for the mental apparatus, seeing as the body's needs are not satisfied through states of wishful thinking.

In the *Interpretation of Dreams* Freud adheres to the idea that a theory of how memory behaves in dreams is of great importance for any theory of memory,¹¹⁹ especially in the sense that it teaches us that nothing, once “mentally possessed”, can be entirely lost, but also insists that dreams do not reproduce experiences (1901:20): “Dreams yield no more than fragments of reproductions; and this is so general a rule that theoretical conclusions may be based on it” (21).¹²⁰ He also comes to the conclusion that dreams are the fulfilments of wishes (121) and that dreams often take the place of action (124). The claim that all dreams are wish-fulfilment dreams, needs to be able to account for the myriad dreams which do not seem to fulfil any wishes, but in actual fact seem anxiety provoking and distressing. Freud accounts for this nature of many dreams by means of an exposition on distortion in dreams, distinguishing between the manifest and the latent content in dreams (135). Accordingly, Freud argues that the manifest content in dreams is often the result of repressed ideas.¹²¹ These undesirable or unacceptable ideas are distorted in an attempt to disguise wish-fulfilment, to put up a defence against a wish (141). Freud believes that censorship and dream-distortion are similarly determined, which gives rise to the suggestion that two psychical

¹¹⁹ In *A Metapsychological Supplement to the Theory of Dreams* Freud places emphasis on the benefit that the study of “normal prototypes” of pathological afflictions, examples of such normal prototypes being states of grief or being in love, dreaming, and sleep (1917:222). He discusses the similarities and dissimilarities that can be detected when comparing dreams with the pathological state of schizophrenia.

¹²⁰ Freud’s description of dream-thoughts has much in common with the post-structuralist conception of the structure of language:

These usually emerge as a complex of thoughts and memories of the most intricate possible structure, with all the attributes of the trains of thought familiar to us in waking life. They are not infrequently trains of thought starting out from more than one centre, though having points of contact. Each train of thought is almost invariably accompanied by its contradictory counterpart, linked with it by antithetical association (1901:312).

¹²¹ Freud believes that similarities which one might experience between typical dreams, fairy tales and other forms of creative writing are rife and not accidental, because the “nature of man” has its origin in the impulses of the mind which are rooted in a childhood that has, in a sense, become prehistoric (1901:46). Elsewhere he explicitly states that symbolism is not peculiar to dreams but “is characteristic of unconscious ideation...and is to be found in folklore, and in popular myths, legends and linguistic idioms, proverbial wisdom and current jokes, to a more complete extent than in dreams (351).

forces are at work in the dream process: one that constructs the wish and one that censors this dream-wish (143-144). Freud speculates that the second agency in this scenario most probably has the ability to determine what becomes conscious and by modifying the material that “passes through” it to consciousness. From this conjecture Freud reaches the conclusion that something becoming conscious is a different act from an idea or “presentation” being formed.¹²² Inevitably, in dreams with distressing content, dream distortion has taken place in order to disguise something that is wished for, but is unacceptable or undesirable in some way. This psychical censorship is also evidenced by the fact that dreams are forgotten, in that the forgetting serves the purpose of repression¹²³ (1901:517). Here, as in the Project, memories are in themselves unconscious and, although having the capacity for becoming conscious, can be as influential while being in an unconscious state (539). The extent of this influence is made explicit in the following paragraph:

What we describe as our ‘character’ is based on the memory-traces of our impressions which have had the greatest effect on us – those of our earliest youth – and are precisely the ones which scarcely ever become conscious (540).

Freud believes that the analysis of dreams can illuminate an aspect of the structure of the mental apparatus. In addition to the psychical agencies thus far believed to be part of dream production – the constructor of the dream wish (*Ucs*) and the censor of the wish (*Pcs*). Here the unconscious has no way to access the conscious, except through the preconscious, which inevitably subjects it to modifications.¹²⁴ When we dream, a regression takes place in the mental apparatus, in other words, excitation moves toward the

¹²² Here Freud describes consciousness as “a sense organ which perceives data that arise elsewhere” (1901:144).

¹²³ Freud is emphatic that dreams are not unique among mental processes and that the retention of dreams in memory is comparable to that of other mental processes (1901:521).

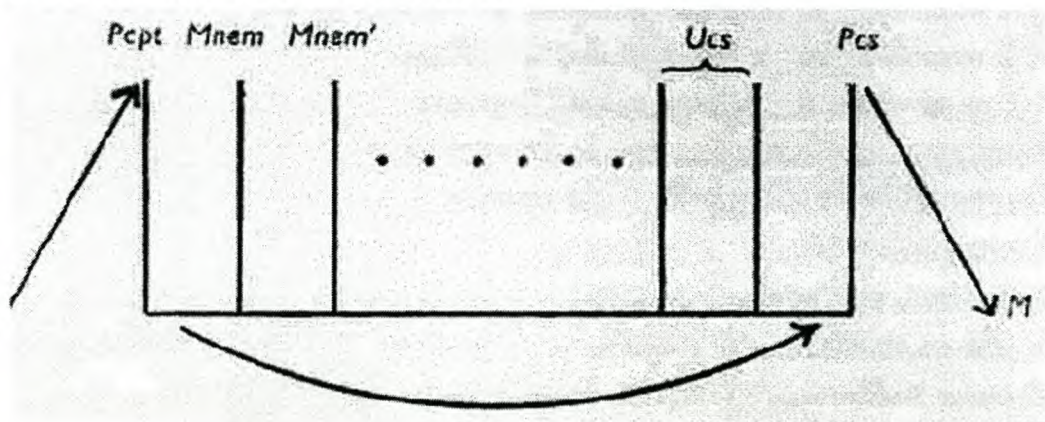
¹²⁴ As we shall see, Freud adds slight alterations to this model of the mental apparatus in *Note upon the “Mystic Writing-Pad”* (1925) and in Chapter VI of *Beyond the Pleasure Principle* (1920).

perceptual system instead of the motor end of the apparatus (or as Freud puts it: “in a dream an idea is turned back into the sensory image from which it was originally derived”).¹²⁵ Regression can also occur in pathological waking states, and particularly involves thoughts that are lined with memories that have been suppressed or have remained completely unconscious (544).

In the *Interpretation of Dreams* Freud confirms the model of the mental apparatus set out in the *Project*, and also concludes that nothing but a wish can set the mental apparatus to work – whether it then leads to hallucinations, dreams or motor action – seeing that the *Ucs* is exclusively aimed at wish-fulfilment (567-568).¹²⁶ He also concludes from his observations that the most complicated thought processes are possible without being assisted by consciousness (593). A train of thought that is “preconscious” might proceed as such without attracting the attention of consciousness or might be suppressed.

Here Freud once again returns to his distinction between the primary and secondary mental processes and elaborates on the scheme (first conceived in the *Project*) of how these processes affect our mental functioning. The primary process (or the *Unc*) is present in the mental apparatus from the beginning, while the secondary process (or the *Pcs*)

¹²⁵ The following scheme (1900:541) constructed by Freud can be useful to clarify his conception of the structure of the mental apparatus:



¹²⁶ In the following section Freud explicitly defines a wish as “the accumulation of excitation...felt as unpleasure and that...sets the apparatus in action with a view to repeating the experience of satisfaction, which involved a diminution of excitation and was felt as pleasure” (1901:598).

develops over the course of time and inhibits the primary process. Because of this belated development of the secondary process, Freud believes that the "core of our being", which would be the "unconscious wishful impulses", exercises a compelling force over mental trends. Furthermore these wishful impulses, which stem from infancy, can neither be destroyed nor inhibited, but *can* be repressed. What the *Interpretation of Dreams* makes clear is that this repressed material continues to exist and remains capable of psychical functioning (608).

Freud is careful to caution his readers against assigning a topographical locality to the mental apparatus. What he has in mind is a dynamic process where a particular mental grouping receives a "cathexis of energy" attached or withdrawn from it, and thus comes under the sway of an agency (610). These systems are not physical entities and are described as being virtual – existing between elements of the organic nervous system "where resistances and facilitations provide the corresponding correlates" (611).

Freud recognises the immense implication that his theories on the unconscious would have on theories of mind.¹²⁷ Whereas, before the advent of the unconscious, "psychical" was considered to be equivalent to "conscious", complex psychical processes now become possible that do not rise to consciousness and, in fact, the unconscious becomes the basis and the larger sphere of mental life. As Freud insists: "It is essential to abandon the overvaluation of the property of becoming conscious before it becomes possible to form any correct view of the origin of what is mental" (612). In *The Ego and the Id* (1923) Freud reiterates his claim in the strongest terms:

The division of the psychical into what is conscious and what is unconscious is the fundamental principle of psycho-analysis; and it alone makes it possible for psycho-analysis to understand the pathological processes in mental life, which are as common as they are important, and to find a place for them

¹²⁷ See the *Unconscious* (Freud 1915c:166-171) where Freud defends his view that the postulation of the unconscious is both necessary and legitimate.

in science. To put it once more, in a different way: psycho-analysis cannot situate the essence of the psychical in consciousness, but is obliged to regard consciousness as a quality of the psychical, which may be present in addition to other qualities or may be absent (13).¹²⁸

In actual fact Freud distinguishes between two kinds of unconscious: the *Ucs* (the content of which is inadmissible to consciousness) and the *Pcs* (the excitations of which, after undergoing censorship, can reach consciousness).¹²⁹ The only role left for consciousness to play is that of “a sense organ for the perception of psychical qualities” (615). The consciousness (*Cs*)¹³⁰ is pictured as a system which is susceptible to excitation by *qualities* but does not have any memory. The qualities that the processes that enter *Cs* from the *Pcs* have attached to them are those of pleasure and unpleasure. Because of this qualitative aspect the *Cs* influences the discharge of the cathexes, which might otherwise only be displaced on grounds of quantities (*Q*). The *Cs* is also capable of creating a new series of qualities, and of a process of regulation, presumably unique to humans: thought-processes are associated with verbal memories, which draw the attention of consciousness to themselves. This characteristic allows for a mobility in the process of thinking.

¹²⁸ At this point Freud adds an admonishment to philosophers which concurs exactly with the sentiments expounded on in this work:

To most people who would have been educated in philosophy the idea of anything psychical which is not also conscious is so inconceivable that it seems to them absurd and refutable by simple logic. I believe this is only because they have not studied the relevant phenomena of hypnosis and dreams, which – quite apart from pathological manifestations – necessitate this view (1923:13).

¹²⁹ Freud is careful to point out that there are two kinds of unconscious in a “descriptive” sense, but that in a “dynamic” sense there is only one (see, for example, 1923:15).

¹³⁰ Later the coherent organisation of mental processes, to which consciousness is attached, will be named the ego. The ego would also be responsible for effecting censorship and repression (1923:17). The *Unc* is later named the *Id* (23). The super-ego, which makes up the third part of Freud’s tripartite division of the mind results from two factors, one biological and one historical. Shortly, the biological origin is that of one’s lengthy state of dependency and helplessness during infancy, and the historical being the gradual suppression of the Oedipus complex under the influence and sanction of parents and society.

Freud picks up this theme again in *Unconscious* (1915c). After examining the evidence suggested by observing schizophrenic patients he concludes that the difference between a conscious and an unconscious presentation is not to do with locality in the brain, but is in fact that “conscious presentation comprises the presentation of the thing plus the presentation of the word belonging to it, while the unconscious presentation is the presentation of the thing alone” (201). Freud explains the process of having a word “added” to such a presentation as follows:

The system *Ucs* contains the thing-cathexes of the objects, the first and true object cathexes; the system *Pcs* comes about by this thing-presentation being hypercathected through being linked with the word-presentations corresponding to it. It is these hypercathexes, we may suppose, that bring about a higher psychical organisation and make it possible for the primary process to be succeeded by the secondary process which is dominant in the *Pcs* (1915c: 202).¹³¹

Being linked with a word does not make a presentation conscious but endows it with the capacity to become so. As Cilliers puts it, Freud presents us with two preconditions for consciousness: language and the unconscious (1989:127).¹³² According to Freud pathological conditions are then the result of the contents conscious and unconscious systems becoming confused or “scrambled”.¹³³

As mentioned before, a major task of *Cs* is to orientate the individual with regard to what is internal and what is external, what is “real” and what is not. Thus *Cs* must have the capacity to do reality testing, which Freud envisions as “having at its disposal a motor innervation which determines whether the perception can be made to disappear or whether it proves resistant” (1915c:233). As we have already seen, *Cs* needs a special faculty,

¹³¹ A presentation that is not put into words remains in a state of repression.

¹³² Freud himself does not elaborate on a theory of language or meaning to illustrate or substantiate this claim.

¹³³ See Freud (1915c:201) for an extensive discussion of this phenomenon

over and above that of detecting pleasure and unpleasure to carry out such reality testing, namely attention. We have also seen that attention is dependent on memory traces of previous experiences, according to which the present sensation can be evaluated. Freud's theory of memory is best illustrated in *Note on the "Mystic Writing Pad"* and the analogy that he develops in this short piece is still compatible with current theories on the development and functioning of memory.¹³⁴

7. A note on the "Mystic Writing-Pad"

Freud takes as his starting point the fact that all people – normal and neurotic – have reason to distrust their memory, and can guarantee its authenticity by some kind of supplement, such as a note (1925:227).¹³⁵ Here we have a permanent memory-trace and we are fairly certain of this "materialised portion of [the] mnemic apparatus" remaining unaltered and undistorted, which is not always the case with actual memory. While this intransience surely appears to be an advantage when it comes to the necessity to recall memories, Freud points out one distinct disadvantage of a permanent record: the capacity for further recordings is rapidly depleted. Not only that, the recorded memory might become obsolete, in which case its indelibility is a distinct disadvantage, since it leads to clutter that will most probably reach unmanageable proportions.

A procedure that will avoid both disadvantages and still provide a material recording of a memory is to use a receptive surface, which can retain its receptivity until one has no need of that particular recording. The ideal would be to be able to remove the recording, without having to discard the entire writing-surface, as when writing with chalk on a slate. The disadvantage

¹³⁴ As stated earlier it is in *Note on the "Mystic Writing-Pad"* that Freud develops a theory of mental functioning without, as Cilliers states it, recourse to ω neurones (1989:116). Cilliers sees this move as vitally important in that it allows for a theory of consciousness as the result of the interaction between memory and perception, rather than postulating consciousness as the property of an ad hoc system (ibid.).

¹³⁵ See Dennett (1991:101-138) for an interesting discussion on how the perspective of the observer is inextricably linked with memories of relevant events, despite the inability of the observer to guarantee the authenticity of these memories.

of such a surface would be that it would be impossible to keep any permanent traces, if one so wished. The slate would have to be wiped clean before any fresh memories can be recorded. Freud concludes that unlimited receptive capacity and the ability to retain permanent traces seem to be impossible in the apparatus that we might use as substitutes to memory (ibid.). The central point that Freud tries to make here is that the faculty of memory is capable of precisely such a function. The memory is capable both of retaining permanent (although alterable) memory traces and of discarding traces that have become redundant. Such an ability is central to the way that memory functions.

Freud finds an apt model for this ability of the mental apparatus in a children's toy – the “Mystic Writing-Pad”. The toy is constructed as follows: the bottom part of the pad consists of a wax-like substance, over this slab is laid a piece of thin, transparent waxed paper, and thirdly a transparent piece of celluloid covers the wax-paper. Both the celluloid and the wax paper are secured to the pad along its top edge and can be lifted from the wax-like substance. To make recording upon this writing-device make use of a pointed stylus and writes on the sheet of celluloid. The stylus presses the surface of the waxed paper onto the wax-like substance of the pad, grooves that are visible as dark lines through the surface of the waxed paper (visible though the celluloid). In order to remove the lines one need only lift the waxed paper and celluloid layers from the wax, breaking the contact between the wax and the layer covering it. The surface of the Mystic Pad is now clear and can receive new recordings. The layer of celluloid acts as a protective sheath over the thin waxed paper, which would be damaged by the stylus. The Mystic Pad differs from paper and chalk in an important way: permanent traces of what has been written are left on the surface of the wax slab and can be read in certain lights (230).

Freud uses the Mystic Pad as an approximation of the structure of the perceptual system of the mind (229).¹³⁶ The celluloid layer is likened to the

¹³⁶ In *Beyond the Pleasure Principle* (1920) Freud contended that the perceptual apparatus of the mind consists of two layers: an external protective shield to dampen the

protective shield against stimuli that protects the *Pcpt.-Cs.* system (the waxed paper) against strong excitations, and that forms no permanent traces. And of course the perceptual system (the unconscious), like the wax pad, retains the imprints or “traces” of preceding stimuli, even while being capable of receiving fresh impressions. The problem posed by the two functions necessary for receiving and retaining memory, without cluttering up the perceptual system, is solved through dividing the functions between two separate, but interrelated systems (230). A major difference between the perceptual system and the Writing Pad is of course that with regard to the perceptual system, memory can be reproduced (brought to consciousness).

On the strength of this analogy Freud returns to an idea that had surfaced in both the *Project* and in *Beyond the Pleasure Principle*:

My theory was that cathectic innervations are sent out and withdrawn in rapid periodic impulses from within into the completely impervious system *Pcpt.-Cs.* So long as that system is cathected in this manner, it receives perceptions (which are accompanied by consciousness) and passes the excitations on to the unconscious mnemonic systems; but as soon as the cathexis is withdrawn, consciousness is extinguished and the function of the system comes to a standstill (231).

To Freud, the periodic break in contact between the wax pad and its two covering layers when the layers are lifted from the pad and the writing disappears, is analogous in the perceptual system to the withdrawal of the cathexis from the unconscious when consciousness ceases to work and becomes “extinguished”. In earlier works Freud also hinted at this discontinuous functioning of the *Pcpt.-Cs.* system is the origin of the concept of time.¹³⁷ Freud concludes his analogy as follows:

strength of the incoming stimuli, and the “surface” or system that receives the stimuli – the *Pcpt.-Cs.*

¹³⁷ See, for example, the Unconscious (1915c:187-188).

If we imagine one hand writing upon the surface of the Mystic Writing-Pad while another periodically raises its covering-sheet from the wax-slab, we shall have a concrete representation of the way in which I try to picture the functioning of the perceptual apparatus of our mind (232).

With the computer becoming commonplace in the second half of the twentieth century, Freud has often been interpreted along cybernetic lines. Cilliers (1989:134-136)¹³⁸ points out that this trend can in some measure be attributed to the desire to link Freud to trends in cognitive science, ostensibly a scientifically respectable successor to behaviourism.¹³⁹ The neurological theory of the Project can be described in terms of information theory, cybernetics, feedback and control (134), and Freud is often represented in these terms. Cilliers cautions against this practice and it is precisely here where he sees the value of the *Note on a "Mystic Writing-Pad"*, and Freud's other metapsychological works.

Erdelyi (Cilliers1989:134) serves as an example of such an attempt to render Freud's model of the mental apparatus in a way that would render it more acceptable to cognitive science. As Cilliers points out, Erdelyi questions Freud's use of the Mystic Writing-Pad as an apt metaphor for the way that memory functions, and attributes Freud's choice of analogy to the fact that the idea of the computer was not available to him. Erdelyi then suggests the calculator as a more apt analogy than the writing-pad. The main advantage, as he sees it, is that the calculator can reproduce its contents – something that the writing-pad cannot do. Erdelyi sees this as an insurmountable shortcoming in the analogy. Freud himself is not unaware of this shortcoming, but would not necessarily have chosen the computer as metaphor, had the idea of it been available to him.

¹³⁸ Also see Dennett (1991:211).

¹³⁹ This view is an apt illustration of contemporary disciplines as the inheritors of many of the presuppositions of the Enlightenment theorists, as discussed in chapter 2.

Cilliers points out a number of shortcomings in Erdelyi's analogy, which take up some of the themes that have been central to this discussion hitherto. The first shortcoming is that the metaphor of arithmetical relations between numbers does not substitute for the metaphor of writing; it is simply not complex enough. Secondly, the relationships between the memory buffers in a computer are axiomatic, while the relationships between the contents of the unconscious are non-rational and *non-causal*. The third shortcoming, according to Cilliers and certainly one that would dismantle Freud's main argument in the Mystic Writing-Pad, is that with the computer, the concept of memory-traces is lost. Storage cells in the computer are isolated from one another, and stored numbers can be recalled in their exact original form. This analogy does not allow for the *interaction* between perception and memory that Freud is at pains to illustrate. Freud wants a metaphor that illustrates that memories are never completely lost, nor are they saved as complete icons in some filing-system of the brain. Through the Mystic Pad he emphasises the influence that traces of memories – the imprints on the wax slab – have on subsequent memories.¹⁴⁰ Finally, the computer metaphor retains the idea of some kind of external operator, a central and executive force that instigates specific actions. This passivity does not reflect the dynamic nature of the mental apparatus.¹⁴¹ In the end, Erdelyi's metaphor is even less successful than the original, which it critiques. The computer metaphor pre-empts many possibilities that may be attributed to the mental apparatus, simply by identifying with specific formal processes (applicable to computers) too quickly (Cilliers 1989:137).

Cilliers proceeds to draw some more links between Freud and neuropsychology, as well as the relationship between the mental processes and language (Cilliers 1989). These links lead us to assert the value of Freud's work to contemporary theory. In fact, as we shall argue in the next

¹⁴⁰ The necessity of his model mimicking the apparatus in that it remains an imprint of what has passed over its surface, is precisely the reason that Freud rejects the chalk and slate metaphor (1925:127-130).

¹⁴¹ See Dennett (1991:101-139) for his refutation of this kind of representation, which he calls the idea of the *Cartesian Theatre*. Also see his reference to the misrepresentation of,

chapter, between Freud and complexity theory we will be able to formulate an extraordinarily viable model of the self. Such a theory might go a long way in filling in some of the gaps left by the overwhelming emphasis placed on the rational in Western theory for the last three-hundred-odd years. At the same time we hope to avoid the threats posed by a reductionistic materialism.

8. Post-Freudian theory

Freud's initially materialist approach to the mind laid the foundation for his dynamic and conflictual theories of the psyche. The mind could no longer be thought of as consisting of all that is rational thought, open to introspection and axiomatic.

Freud delegated what has traditionally been considered to be mind to the realm of the ego. The ego is wholly conscious and is also and is responsible for learning and adapting to the external world (physical or social environment). While the ego is concerned with perception, memory, speech, etc. it is not a self-contained entity. The mental apparatus also consists of the id and the super-ego, both unconscious, but with a profound influence on the ego. The super-ego as the instrument of conscience and suppression, and the id, which operates according to the demands of the instincts and not those of logic or external reality, make for a tumultuous mind. The mind/ego of theories of mind up until this point has become a property of a much more intricate system and process than had ever been conceived of before (Zangwill1998b:278). After Freud, many presuppositions that had been taken for granted in the rationalist/empiricist debate had to be re-evaluated, and an isolationist theory of mind became all the more implausible for it.

The dichotomy between mind and brain decreased in some circles, but increased in others, especially due to the boom in the neurological sciences, which on the whole steered clear of the issues of consciousness and the self. Sperry (1998:164) asserts that as the apparent dichotomy between mind and

what he calls, *Von Neumann machines* (i.e. computers) in the press, practically from their first conception (1991:214).

brain became greater, the more successful the neurosciences became in explaining brain activity in terms of its electrophysiology, chemistry and anatomy. The idea that the subjective inner experiences of the subject, its mental states, could influence brain function became more implausible than ever. The 1960's saw extreme materialist positions, such as that of Armstrong, where the mind is nothing but the brain, and less extreme versions of materialism, where subjective (mental) phenomena could still have a causal impact on activities of the brain.

Sperry adhered to the latter and summarises this theory in a paragraph of such denseness and economy that it serves our purposes better to copy it *verbatim*:

The neural infrastructure of any brain process mediating conscious awareness is composed of elements within elements and forces within forces, ranging from subnuclear and subatomic particles at the lower levels upward through molecular, cellular, and simple-to-complex neural systems. At each level of the hierarchy, elements are bound and controlled by the enveloping organizational properties of the larger systems in which they are embedded. Holistic system properties at each level of organization have their own causal regulatory roles, interacting at their own level and also exerting downward control over their components, as well as determining the properties of the system in which they are embedded. It is postulated that at higher levels in the brain these emergent system properties include the phenomena of inner experience as high-order emergents in the brain's hierarchy of controls...Interpreted as holistic high-level dynamic properties, the mental phenomena are conceived to control their biophysical, molecular, atomic, and other sub-elements in the same way that the organism as a whole controls the course and fate of its separate organs and cells, or just as the molecule as an entity carries all its component atoms, electrons and other,

subatomic and subnuclear parts through a distinctive time-space course in a chemical reaction (1998: 164-165).

This *interactionist* theory allows for interaction between neurophysiology and mental events. These mental events are in part determined by their neural components, but they are also determined by the “spacing and timing” of these components. The space-time properties of the neuronal infrastructure is added to causal accounts of the brain/mind’s activities. In discussions on interactionist models, much is made of the conception of mental events as *supervening*, rather than *intervening* in the physiological process. Mind directs neuronal events, without interacting with the components of the brain, just as “an organism might move and govern the space-time course of its atoms and tissues, without interacting with them (165). Consciousness now has a causal role in brain function. This principle of top-down emergent control, or emergent determinism, applies, according to Sperry, to all hierarchic systems in all science (ibid.).

Here we are presented with a marriage between metaphysical and material theories of mind – a third possible philosophical position, namely mentalism. It will be our position that this is an uneasy marriage. Sperry insists that *mental* factors¹⁴² retain the possibility of overriding the “subsidiary” forces of the neural substructure of the brain (165). In a telling phrase he declares that “the mind has been restored to the brain of experimental science” (166). Sperry’s approach is much closer to a theory of mind and self that this study would endorse. What we will propose here as an amendment to the approach of Sperry is dropping the idea of “mental events” altogether, and to promulgate a moderate materialism.

¹⁴² Sperry (1998:165) gives as examples of mental events one’s personal wishes, feelings and willed choice.

9. Dennett's materialism and the key to a demystified "mind"

In *Consciousness Explained* (1991) Dennett differentiates between different phenomena, which make up what we call "consciousness", and insists that all of these phenomena are the result of the physical activities of the brain (16). He insists that the mind *is* the brain and that all instances of "evidence" to the contrary are illusions created by the properties of these processes. Dennett's materialism will underpin our effort to demystify the self, and to present it as one of the said processes of consciousness. All that will remain for us to do then will be to develop a theory of the self, based on the principles of complexity theory.

Dennett situates the origin of the concept of self with a peculiarity particular to conscious events: they are "witnessed" or experienced. An experience has to be *somebody's* experience; someone must think it, or feel it, or imagine it. And at first blush, brains do not seem to be anything akin to what we would imagine such an *experiencer* to look like. Hence, the idea of a self (or soul, or ego, or person) as distinct from a brain or a body. In contrast to Kenny, though, Dennett does not put confusion over the self down to grammatical error (29). He does, however, propose that a scientific study of mental phenomena needs to be conducted in the third-person perspective.¹⁴³ The supposed impossibility of conducting studies on consciousness from a third-person perspective is of course the main objection that theorists like Nagel and Searle have against the possibility of a scientific study of consciousness.¹⁴⁴

Dennett's method of *heterophenomenology* extracts texts from the speaking subject and uses those texts to generate a *theorist's fiction*, the "heterophenomonological" world of the given subject. This fiction is an account of all that the subject sincerely believes to exist in his/her conscious experience. Dennett insists that such a narrative *is* a portrayal of exactly what

¹⁴³ See Dennett's discussion of his proposed method, namely *heterophenomenology* (1991:73-98).

¹⁴⁴ Cf. "What is it like to be a bat?" (Nagel 1982:391-403) and "Minds, Brains, Programs" (Searle 1982:353-373).

it is like to be that subject, and is an adequate basis from which to explore this heterophenomenology:

The heterophenomenology exists – just as uncontroversially as novels and other fictions exist. People do undoubtedly believe they have mental images, pains, perceptual experiences and all the rest, and *these* facts – the facts about what people believe, and report when they express their beliefs – are phenomena any scientific theory of the mind must account for. We organise our data regarding these phenomena into theorist's fictions, “intentional objects” in heterophenomenological worlds. Then the question of whether items thus portrayed exist as real objects, events, and states in the brain – or in the soul, for that matter – is an empirical matter to investigate (1991:98).

It seems that Dennett has paved the way for an empirical theory of mind.

Dennett would agree with both Nagel and Searle that a conscious mind is an observer and that where there is a mind, there is a *point of view* (101). The logical implications obvious to these theorists in this simple assumption breaks down for Dennett, however, when he tries to pinpoint a point of consciousness within a brain or an individual. Dennett insists (rightly in our view) that the brain is “headquarters” of the perceived observer, and that there is no other, deeper, headquarters in the brain, where consciousness is seated. Dennett calls the view that some such central observer exists within the brain as *Cartesian materialism*.¹⁴⁵ The Cartesian pineal gland would be a candidate for such a “Cartesian Theatre” (107). Dennett’s objection against the Cartesian Theatre is concisely summed up in the following quote:

¹⁴⁵ For absolute clarity on what Dennett means with this term we quote him verbatim:

Cartesian materialism is the view that there is a crucial finish line or boundary somewhere in the brain, marking the place where the order of arrival equals the order of “presentation” in experience because *what happens there* is what you are conscious of (1991:107).

The Cartesian Theatre may be a comforting image because it preserves the reality/appearance distinction at the heart of human subjectivity, but as well as being scientifically unmotivated, this is metaphysically dubious, because it creates the bizarre category of the objectively subjective – the way things actually, objectively seem to you even if they don't seem that way to you! (Smullyan 1981, quoted in Dennett 1991:132).

A brain without a *Cartesian Theatre* significantly complicates the concept of the point of view of the subject. The observer's subjective sense of sequence must then be determined by something other than "order of arrival" of experienced items to the locus of consciousness¹⁴⁶ (ibid.). As already discussed, Freud, by lighting upon the idea that a vast number of our "mental states" or mental processes belong to the realm of the unconscious, threw the proverbial cat among the pigeons in terms of established presuppositions of what the mind is. As discussed in the previous section, not only are a vast number of our mental processes not conscious, but it becomes possible for the subject to deny the existence of mental states, which are unconscious, but nevertheless active in his/her mental processes. Towards the end of the twentieth century, and the beginning of the twenty-first, the existence of the unconscious is not questioned, but how *consciousness* comes about is a major point of contention¹⁴⁷. Freud's theory did not dispose of the Cartesian Theatre in Dennett's sense, for him, consciousness occurred when presentations reached a certain part of the mental apparatus – the Cs.

Dennett proposes the *Multiple Drafts* version of how consciousness comes about in the place of the concept of the *Cartesian Theatre* (101-170). The Multiple Drafts model asserts that, in his words: "all varieties of perception – indeed, all varieties of thought of mental activity – are accomplished in the brain by parallel, multitrack processes of interpretation

¹⁴⁶ See Dennett's discussions on the unfeasibility of theories of central or conscious "observers" in the brain (1991:101-111 and 126-134).

¹⁴⁷ Recall that Freud believed the possibility of consciousness to come about once a mental presentation is put into words, or linked with language.

and elaboration of sensory inputs" (111). Or, as he also has it, information entering the nervous system is under continuous "editorial revision".¹⁴⁸ In fact, he argues that it is misleading to ask the question of when perceptions become conscious. The information content gleaned from sensory inputs is distributed throughout different systems of the brain. These, what he calls "distributed content-discriminations" become something like a narrative stream – a multiplicity subject to continual editing by many processes distributed in the brain. "[A]t any point in time there are multiple "drafts" of narrative fragments at various stages of editing in various places in the brain" (113).¹⁴⁹ There is no single, final narrative, which is delivered to consciousness and can be considered to be the actual stream of consciousness of the subject. In other words, there is no point in the brain where it all comes together. As Dennett notes, some contentful (sic.) states might die out completely, leaving no trace, while others do leave traces that might later arise in some form or another, for example a verbal report, or an emotional state (135).

Having done away with the Cartesian Theatre and expounding on the merits of the Multiple Drafts model, Dennett is free to explore what implications his model will have for our conceptions of mental functioning. Many assumptions instituted in philosophy of mind now need to be re-evaluated. One definite advantage of the Multiple Drafts model is that it lends itself to a theory of the evolution of consciousness. Far from being a metaphysical, non-bodily phenomenon, here we have a picture of consciousness that developed with a species, presumably in accordance with constraints and possibilities imposed by the environment and genetic adaptations (171-226)¹⁵⁰. Approaching consciousness from an evolutionary

¹⁴⁸ See Dennett 1991: 111-138 for discussions of different psychological experiments that support such a conception of editorial revision.

¹⁴⁹ See Dennett's discussion (1991:115-126) on whether these revisions are "Stalinesque" or "Orwellian."

¹⁵⁰ See Dawkins (1976:62-64) where he develops the idea that consciousness evolved because having the capacity to simulate scenarios and experiment with possible outcomes (i.e. being conscious in the sense of factoring a model of oneself into perceptions of the environment) gives an organism a competitive edge over organisms lacking this ability.

perspective, also allows for the possibility that consciousness is not static, nor is it necessarily optimal in its present state¹⁵¹.

Dennett presents the design of human conscious minds as the result of three evolutionary processes (173). The need for self-preservation and control through the ability to track and anticipate, gave rise to the nervous system in successive guises. Systems proficient in information gathering, geared towards information that is beneficial to the organism, develop. In other words, these systems become part of the innate design of the nervous system. These states are not necessarily conscious states.¹⁵² The development of nervous systems that have an element of plasticity and therefore have the ability to learn in the course of their lifetime provided other (other than genetic) “mediums” for evolution of the nervous system to occur and hence speed up the process hitherto driven by natural selection and genetic mutation (182). Such a learning mechanism would operate along the same lines as “natural” evolution, in other words, a process of evolution through selection.¹⁵³ Dennett refers to this process as *post-natal design fixing* (183) – hence, it could roughly be characterised as a process of learning rather than development.¹⁵⁴ Plasticity allows the brain to reorganise itself in some ways, and so adapt to its environment. A plastic, adaptable brain (the cortex) is the first “new” medium in which the evolutionary process with regard to nervous systems can be speeded up. In fact, Dennett attributes the radical transformation of human society in the last 10,000 years (the development of agriculture, art, cooking, etc.) to new ways our ancestors developed of harnessing mental capabilities. He uses the metaphor of creating software, which could be run on the wired in hardware of the *homo sapiens* brain (190).¹⁵⁵

¹⁵¹ As we shall see in the final chapter, the self also lends itself to evolutionary approach, and we will attempt such an approach with the help of Richard Dawkins and his theory of *memes*.

¹⁵² See Dennett 1991:171-193 for a more detailed discussion.

¹⁵³ There are many and varied theories how such a process would work, but constrained space does not allow a discussion of this interesting issue here.

¹⁵⁴ This distinction is by no means clear, but for simplicity's sake we shall refer to post-natal design fixing as “learning”.

¹⁵⁵ The reader will remember that the computer-metaphor was discredited in an earlier section of this chapter. Dennett's software-metaphor, however, will be exempt from the

The second new type of evolution that Dennett discusses, is *cultural* evolution, and the transmission of its products to others (193). This evolutionary medium is the product of the plasticity of the brain, which makes learning possible. Through cultural transmission we install developed “programmes” of behaviour in developing (usually young) minds.¹⁵⁶ Dennett calls this process of relating information “software-sharing”, which happens, of course, through some or other form of language (194). Through honing the art of software-sharing, culture develops into what Dennett calls “a repository and transmission medium for innovations” (199).¹⁵⁷ And so, as with Freud, Dennett progresses to the importance that culture has for the existence and development of consciousness.¹⁵⁸ One of the first steps in the process of self-design that human engages in after birth, is to acquire language. One could go so far as to say that, prior to language, the self, in any meaningful sense of

similar criticism, because of its usefulness for the present discussion. Dennett does allow for the restrictions inherent in his metaphor, emphasising that the computer (or the “von Neumann-machine”, at least) consists of serial architecture while the brain is a parallel processing machine (1991:215-217 and 219-222). He elaborates on his metaphor as follows:

A computer has a basic *fixed* or *hard-wired* architecture, but with huge amounts of plasticity thanks to memory, which can store both programs (otherwise known as software) and data, the merely transient patterns that are made to trace whatever it is that is to be represented. Computers, like brains, are thus incompletely designed at birth, with flexibility that can be used as a medium to create more specifically disciplined architectures, special-purpose machines, each with a striking individual way of taking in the environment’s stimulation (via the keyboard or other input devices) and eventually yielding responses (via the CRT screen or other output devices) (1991:211).

The plasticity within the computer makes “virtual machines” possible, thus different patterns imposed on the hard-wired machine will lead it to perform different functions and hence creates different possible *virtual* machines.

¹⁵⁶ It is important to note that this type of “programming” is firmly rooted in the material mechanisms of the brain, as is evidenced by the following statement by Colwyn Trevarthen:

...there is increasing evidence that the self-organizing processes of brain tissue formation continue to have a hand in even the most specialized and culturally elaborated acquisitions of learning. There are regions of the cerebral cortex in the foetus that appear to be specially formed to engage in cultural life and acquire traditional skills (1998:107).

¹⁵⁷ Or, as Trevarthen has it: “Growing human brains require cultivation by intimate communication with older human brains” (1998:108).

¹⁵⁸ The final chapter will be devoted to the role that culture plays in the development of the self, especially with reference to Dawkin’s concept of memes.

the word, does not exist¹⁵⁹. The capacity for designing and developing a self does, however, exists.

Dennett sums up his argument in the following paragraph, which will also serve very well as bridge to the next chapter where discuss complexity theory and its relevance to the process of developing a self:

Human consciousness is *itself* a huge complex of memes¹⁶⁰ (or more exactly, meme-effects in the brain) that can best be understood as the operation of a “*von Neumannesque*” virtual machine¹⁶¹, *implemented* in the *parallel architecture* of a brain that was not designed for any such activities. The powers of this *virtual machine* vastly enhance the underlying powers of the organic *hardware* on which it runs, but at the same time many of its most curious features and especially its limitations, can be explained as byproducts of the *kludges* that make possible this curious but effective re-use of an existing organ for novel purposes (210) (cf. footnote 63 and 65).

This paragraph will guide the discussion of the final two chapters, at the end of which its relevance to a complexity-theory approach to the same questions will hopefully be clear. All that remains to be touched upon with regard to the current discussion is the final conclusion that Dennett draws in his discussion on the evolution of consciousness. He believes that

¹⁵⁹ Dennett places much emphasis on the role that writing plays in structuring consciousness:

...[N]ot just spoken language, but writing plays a major role, I suspect, in the development and elaboration of the virtual machines that most of us run most of the time in our brains...the virtual machine...can only exist in the environment that has not just language and social interaction, but writing and diagramming as well, simply because the demands on memory and pattern recognition for its implementation require the brain to “off-load” some of its memories into buffers in the environment (1991:220).

¹⁶⁰ Memes are the cultural equivalent of genes. Dawkins’s conception of memes will be discussed in detail in the final chapter.

¹⁶¹ I.e. a computer with a fixed (hardware) structure that can run different kinds of soft-ware, and as such can function as a series of different “machines” with divergent capabilities.

consciousness arose because the brain “[had] to become the object of its own perceptual systems” (1991:225). As we shall see in the following chapter, with this assertion Dennett reinstates the qualitative, colourful and value-rich world of inner experience, long excluded from the domain of science by the behaviourist-materialist doctrine.

Chapter Four

The Complex Self

Looking on the bright side, let us remind ourselves of what has happened in the wake of earlier demystifications. We find no diminution of wonder; on the contrary, we find deeper beauties and more dazzling visions of the complexity of the universe than the protectors of mystery ever conceived. The “magic” of earlier versions was, for the most part, a cover-up for frank failures of imagination, a boring dodge enshrined in the concept of a *deus ex machina*. Fiery gods driving golden chariots across the skies are simpleminded comic-book fare compared to the ravishing strangeness of contemporary cosmology, and the recursive intricacies of the reproductive machinery of DNA make *élan vital* about as interesting as Superman’s dread kryptonite. When we understand consciousness – when there is no more mystery – consciousness will be different, but there will still be beauty, and more room than ever for awe

Dennett (1991:25)

The importance placed on [the] relationship [of the brain to language] should not lead to the conclusion that the brain is something that operates *on* or *with* language like a kind of word processor. The contents of the brain are not propositions, attitudes, beliefs, statements, intentions, or whatever linguistic entities you wish. The brain is *like* language, the structure and the functioning of language. “Language” should also be seen in the general sense of a system of symbols that enable communication, whether pictorial, hieroglyphic, graphic or auditory, and not as any specific natural language. The detailed functioning of two brains may be as different as the difference between Swedish and Swahili, and as similar as the similarities.

Cilliers (1989:49)

1. Introduction

By way of recapitulation the following paragraph of Dennett’s should concisely sum up his conclusions on the evolution of consciousness as discussed in the previous chapter:

In our brains there is a cobbled-together collection of specialist brain circuits, which, thanks to a family of habits inculcated partly by culture and partly by individual self-exploration, conspire together to produce a more or less orderly, more or less effective, more or less well-designed virtual machine... By yoking these independently evolved specialist organs together in common cause, and thereby giving their union vastly enhanced powers, this virtual machine, this

software of the brain, performs a sort of internal political miracle: It creates a *virtual captain* of the crew, without elevating any one of them to long-term dictatorial power. Who's in charge? First one coalition and then another, shifting in ways that are not chaotic thanks to good meta-habits that tend to entrain coherent, purposeful sequences rather than an interminable helter-skelter power grab (1991:228)¹⁶².

In keeping with Dennett's materialist approach to the mind, and his emphasis on how evolution, both cultural and genetic, contribute to the structure and functioning of the brain, we move to Paul Cilliers's contention that the brain can be considered to be a complex system. As with Dennett's argument, this materialist approach will lead to the conclusion that the higher functions of the brain are grounded in the physiology of the brain. More specifically, that consciousness can be explained as an *emergent property* of the brain. The basic tenets of complexity theory as discussed in chapter 3 can readily be recognised in Cilliers's discussion of the physiology of the brain and in his conclusion that the structure of the brain lends itself to be modelled as a complex, distributed system (cf. Cilliers 1989:57-104). The brain is interesting with regard to complexity theory, not only because of its complex structure, but also because of its ability to deal with complexity – learning about and performing complex tasks, for instance (1998:16). The brain is a complex system functioning within a vastly complex environment. In fact, as we shall see, differentiating between brain, as such, and its environment becomes all the more difficult, the more we learn about the structure and the development of the brain and brain-processes.

¹⁶² Cf. Trevarthen (1998 :101-110) where he discusses the development of the brain from the embryonic stage through to the mature brain. From his discussion it becomes clear that the brain does have a certain "innate" structure that develops without the benefit of experience (i.e. in the womb). But, while the anatomy of the brain is remarkably complete at birth (roughly a two-thirds-sized likeness of the adult structure), the brain is still far from complete. Post-natally astronomical growth occurs where the nerve-cells form more (and more effective) connections with other cells. This period of growth seems to largely rely on stimuli, which are actively sought and taken up by the baby (103-104). See also Restak (2001) where he discusses how brain growth continues over the entire life-time of the human

Cilliers (1989:8-9) insists that theories of the workings of the brain (whether it be on a neurological or a psychological level) cannot describe the brain as a formal, closed, system. On the contrary, he would eventually contend that, on an anatomical level, neuronal and subneuronal activities lend themselves to be modelled as complex systems (1998:16-18). On a functional level, the brain consists of a vast network of interconnected neurones. He reiterates the importance of taking the available empirical information into account, even when engaging in an ostensibly “philosophical” study of brain and consciousness as follows:

An understanding of how the brain (and specifically the cortex) goes about its task is more important than one would think. To know how (and how accurately, if at all) the outside world is represented in the cortex, how external impulses interact with and cause certain behaviour, emotions, and the notions of “will” and “self”, how we acquire skills and habits, what happens when the brain malfunctions, what memory could be and how language works, we must have a better idea of how the cortex actually works. The study of psychological epiphenomena is certainly useful, but leaves a deeper layer of relations covered. We know something about neurones and something about psychology, and the gap in between is either ignored (by behaviourists) or talked away (by linguistic dualists). But in this very gap lies the enigma of consciousness and self-consciousness (1989:72).¹⁶³

The reader will recall that in the previous chapter Dennett made a similar assertion and concentrated especially on the evolutionary aspect of the cerebral cortex in order to explain the development of consciousness.

being. Crudely simplified, it seems that genes and stimuli interact and as a result cause rival, possible adaptive alternatives to be realised, or not realised, whatever the case may be.

¹⁶³ Note how this assertion mirrors the sentiments of Freud, as discussed in the previous paragraph. The belief that psychology can be grounded on neurological principles (Zangwill 1998b:277) is what led Freud to write his *Project on a Scientific Psychology*. As has

Cilliers does not study the origin or evolution of consciousness, but focuses on the structure and functioning of the nervous system to explain the nature of consciousness. The importance of his work to our discussion essentially lies in taking it as an example of how a complexity theory-based model of the physiological aspect of the brain can provide a plausible and useful account of the “higher functions” of the brain, in this case consciousness. The aim is then to use the same strategy to develop a similar model for the self.

2. Some of the complexities of brain structure in a nutshell

The central nervous systems of all organisms that possess one are geared towards receiving and responding to information. The central nervous system of vertebrates consists of the spinal cord and the brain (Nathan 1998:515) and possesses both inherited structure – genetically determined ways of behaving (Dennett’s hard-wired structure) – and the capability to change some of this behaviour.¹⁶⁴ As we shall see, environmental influences play a decisive role in making changes to the inherent structure of the nervous system.

The major divisions of the adult nervous system read as follows (from top down): the cerebrum, the midbrain, the thalamus and hypothalamus, the cerebellum and the spinal cord. The brain is divided into two hemispheres, which are connected through the corpus callosum (Restak 2001:2). The development of the brain commences early on in the foetal stage, and one of the last areas in the brain to develop is that of the cerebral hemispheres. During the last two months of gestation and during the first few months of infancy this development is at its peak (Trevvarthen 1998:107; Restak 2001:2-34; 37-44). The most significant feature of the cerebrum is its thin outer layer, the cerebral cortex, which makes up seven-tenths of the adult human nervous system (Restak 2001:5).

already been mentioned, Freud never relinquished the belief that psychology could become a “science” comparable to the traditional physical sciences.

¹⁶⁴ The nervous system spontaneously becomes active early *in utero* (Nathan 1998:517).

According to Trevarthen, this rapid development is due to neurones branching out and forming connections that “integrate powerful cortical integrating tissues with the rest of the brain to make conscious perception, voluntary action, and intelligent learning possible” (1998:107). The youngest cells that are situated towards the outside of the hemispheres will only reach maturity towards adolescence. During the first few months of infancy redundant axons are eliminated so that, in the mature brain, the cortex consists of columnar territories (uniform in size), interconnections that link its parts, and structures deeper within the brain. After four months the *corpus callosum* gains in bulk, and fine-tunes communication between the two hemispheres of the brain. In the mature brain information about perceptions, memories and fine motor co-ordinations pass between the hemispheres through the *corpus callosum*.

Trevarthen makes it very clear that the anatomy and function of the cerebral cortices are variable and that patterns in mental abilities between people are the result of their brains growing in different forms (109). As he so vividly puts it: “Males tend to differ from females, left-handers from right-handers and architects from psychologists” (*ibid.*). Some of the diversity of human minds is genetically pre-programmed and is evidenced in brain-tissue development, *but*, Trevarthen insists, the same processes will be influenced by stimuli from the intrauterine and the external environments as well (*ibid.*)

The cerebral cortex is a late evolutionary development, and especially well-developed in man. The cerebral hemispheres are concerned with the activities usually categorised as mental: problem-solving, remembering, planning, imagining, making judgements, forming opinions, etc. The cerebral hemispheres are then also the regions of the brain that can differ overtly from person to person (Nathan 1998:531). The temporal lobes in the cerebral hemispheres are the main areas for memory, while the limbic system mainly organises the essential drives, instigated by the emotions. The hypothalamus has many functions, including: organising hormonal control, circadian rhythms, food and drink intakes, excretion, organising sleep and wakefulness, states of aggression or timidity. The rostral parts of the frontal lobes and their connections to the thalamus, the hypothalamus, and the septal areas are the

regions of the brain most concerned with social behaviour (*ibid.*), and the maturation of these areas continues until puberty (Nathan 1998:531; Restak 2001:71-76). The frontal lobes play an important part in determining energy and are concerned with mood, and inherited and acquired social behaviour.

The nervous system is essentially composed of nerve cells or neurones.¹⁶⁵ Neurones are single cells, and consist of a cell body, the dendrites that convey impulses to the cell body and the axon that relays impulses from the cell body.¹⁶⁶ The axon with its surrounding membrane is called the nerve fibre and the point where it ends on another neurone is called the nerve end. The nerve fibre connects to other neurones in the synapses. Neurones, unlike other cells, cannot divide and reproduce, which means that any neurones that are lost are irreplaceable. Receptors are connected to the central nervous system and convert the energy that they receive (from the senses, for example) into electric current, which is then passed on to the neurones.¹⁶⁷

With regard to our discussion on consciousness it is interesting to note that not all information recorded by the receptors is passed onto the higher levels of the central nervous system; processing of the data already begins at the sense organ (cf. Nathan 1998:515). In a similar vein, all the information reported by receptors does not necessarily reach consciousness (*ibid.*).

The nerve fibres that connect neurones pass nerve impulses between neurones. Each neurone can be likened to a “processor” (analogous to the light-bulbs in the Boolean network discussed in Chapter 2) that calculates the sum of its inputs, and then, if this sum succeeds a certain threshold, generates an output (1989:61). As with the light-bulb example, the output of

¹⁶⁵ The standard estimation of the number of neurones that comprise the human brain is usually around 10^{12} (Gaze and Taylor 1998:543).

¹⁶⁶

¹⁶⁷ Receptors are classified as exteroceptors that report events in the outside world, and interoceptors, that report the internal states of the body. Again current theory was foreshadowed by Freud’s model of the nervous system (cf. pp. 101-105).

this particular neurone in its turn becomes the input of all the neurones that are connected to it. Not only can neurones be connected to a vast number of other neurones at the same time (whether directly or indirectly), they can also be connected to themselves, usually with other neurones as intermediaries.

The connections between neurones are mediated by synapses (a minute gap between the nerve endings of one neuron and the cell body of the next), which regulate the strength of incoming signals and can determine whether the signal would excite or inhibit the neurone. This arrangement allows for flexibility and is one of the keys behind the plasticity of the brain (Restak 2001:9-10). A chemical substance, called a neurotransmitter, is put out into the synaptic gap when an impulse reaches the end of a nerve fibre, and allows the impulse to pass the gap (if this sum exceeds a certain threshold) (Nathan 1998:518).¹⁶⁸ The connection between any two neurones has a certain “weight” which determines the strength of the influence of these neurons on one another (Cilliers 1998:16). The characteristics of the neuronal network seem to be determined by the values of these weights.

In Chapter 2 much emphasis was placed on the ability of complex, “living” systems to be able to self-organise. The brain is no exception. In order to be able to give an adequate explanation of the workings of the brain, without the help of some kind of external or internal controller that supervises or programmes actions, the firing of the neurones and the value of the weights between them need to be *self-organising* somehow. The organism needs to be able to learn from experience and this acquired information has to be incorporated into, and has to influence the workings of the brain.

To be able to learn, we need physical records in the brain – changes need to be brought about in the structure of the brain.¹⁶⁹ The problem of memory is finding the mechanism that establishes this change. Young’s

¹⁶⁸ This is inevitably a vastly simplified description. See Nathan 1998:514-534 for a concise account of how the nervous system functions, in terms of its neuronal composition.

¹⁶⁹ To quote Cilliers in this regard:

The way in which the brain carries and uses the traces of our personal and cultural histories, integrates it with new experiences and maintains a balance between the self and the not-self, the other, is not merely a way of coping with a conscious world, it is the very basis of consciousness (1989).

(1998:455) physiological account of the mechanisms of memory begins with hereditary genetic dispositions. The initial basis of memory in the nervous system is provided by “genetic memory”, which establishes the neuronal pathways of the foetus during gestation. After birth this original multiplicity of possible actions is then “fine-tuned” through interactions with the environment, and redundant neuronal pathways are pruned away, while others are strengthened.¹⁷⁰ In accordance with Dennett’s stipulation as discussed in the previous chapter – that we should do away with the Cartesian idea of a central executive agent within the nervous system – the workings of the system of neurones must account for the formation and strengthening of some pathways and not others. Cilliers discusses Donald Hebb’s famous use-principle in this regard (17). According to Hebb’s principle the connection strength of a synapse between two neurones should increase proportionally to how often it is used (*ibid.*).¹⁷¹ The stronger pathways’ synapses become more effective and are used more often, while unused pathways wither away (Young 1998:455). In this way the mass of largely undifferentiated neurones can develop a structure that is based on the information available to each neurone locally. In other words, the networks of neurones learn from inputs available to them and accordingly develop a (genetically constrained) structure.

Trevarthen (1998:102) places much emphasis on the interaction between genetic predisposition and environmental stimuli in the development of the brain, including the development of the “higher psychological processes,” and declares that after contemporary research done on this subject, the way in which we think about consciousness and human understanding can never be the same (i.e. Cartesian). Although, the brain lays the foundations of even these processes before birth, Trevarthen insists

¹⁷⁰ As discussed in the preceding chapter, Freud had already developed this basic scheme in his *Project for a Scientific Psychology*. He would later on develop this basic mechanism of the brain to such an extent that internal and external stimuli and the primary caretaker become central to the process of the infant’s development of consciousness. Not only does Freud foreshadow contemporary neurology, his work provides the first step in linking consciousness (and the self) to the body: the unconscious.

¹⁷¹ To paraphrase Young: The majority of neuroscientists seem to believe that memory depends on synaptic change (Young 1998:455).

that viewing the brain as pre-wired, with the possibility of affecting changes to the pre-wired system after birth through conditioning,¹⁷² is no longer feasible. Such theories, in his opinion, do not take the actual development of the brain into account.¹⁷³ Evidence suggests that that interacting nerve cells make up and co-ordinate basic rules of perception, prior to birth, in other words without the benefit of experience. As has already been mentioned development in the brain continues after birth, through infancy and into puberty and through into adulthood. This development is dependent on *both* genes and stimuli:

After birth, stimuli are sought and actively taken up by a baby, not just submitted to. Those stimuli which are assimilated cause selections to be made from among rival adaptive alternatives within general adaptive rules for brain formation. These ground rules, including rules for recognizing other persons and for detecting their emotions from their expressions, are innate in the sense that they are formulated in earlier stages when stimuli had no effect. The learning involved takes place as part of a most elaborate developmental strategy that must be ascribed to a continuous regulated unfolding of nerve-cell interactions from the embryo to adult (104).

When the nervous system is formed, nerve-cells form into patterned aggregates and make up patterned circuits. These then communicate biochemically. Once the network of nerve-connections is formed, communication can also take place through the conduction of nerve impulses. These impulses can cause adjustments to be made in the biochemistry of the nerve-cells (cf. Trevarthen 1998:104). A sort of editing process of connections takes place and many connections formed in the embryonic period are removed (Trevarthen 1998:106; Restak 2001:11-19).¹⁷⁴ After birth a whole new array of stimuli become available to be experienced, and cause more

¹⁷² This is of course Pavlov's famous theory of conditioning.

¹⁷³ Cf. Trevarthen (1998:102) for a lengthy exposition on this claim.

¹⁷⁴ The process of selecting desirable connections and eliminating redundant ones is also called *pruning* (Restak 2001:18).

changes in the cellular structure of the cortex than occurs at any other time. Psychological representations and cognitive programmes are elaborated on. Trevarthen notes that the mechanisms responsible for brain growth that operate before birth and in infancy, are continuous with the mechanism for learning from experience throughout life (107).¹⁷⁵ After birth, brain development depends on how stimuli from the outside world interact with the inner states of the brain. In Trevarthen's apt phrasing: "Growing human brains require cultivation by intimate communication with older human brains" (108).

The system of neurones, despite a very limited repertoire of possible interactions, is capable of performing extremely complex tasks (Nathan 1998:518). This echoes our stipulation in chapter 1 that a complex system consists of many components (nodes, neurones, etc.) that interact in such a manner as to produce emergent properties (behaviour) that cannot be reduced to the sum of the parts of the system.¹⁷⁶ Or to quote Cilliers: "Complex behaviour emerges from the interaction between many simple processors that respond in a non-linear fashion to local information" (1998:18). This is of course a very crude simplification of the extremely complex structure and functioning of the brain. As Cilliers (1989:63) notes, not only do the vast amounts of components involved, and the immensely complex physical structure of neurones complicate modelling the functioning of the brain, but we don't fully understand the way in which neurones work.

It seems that a paradox of the brain, is that it is a structurally highly organised, but functionally highly dispersed – in fact, when it comes to the functioning of the brain structure and function cannot be separated (Cilliers 1989:69; Luria 1998:489-490). Cilliers is careful to emphasise that we cannot attribute sole responsibility for certain functions to certain areas of the brain. This is because information is integrated and dispersed throughout the brain. He does not deny that certain areas can be identified with certain primary

¹⁷⁵ It seems that the prefrontal lobes do not reach full maturity until a person's 20's (some speculate that maturity is reached even later) (Restak 2001:76).

¹⁷⁶ This calls to mind Holland's analogy of an ant colony that performs complex tasks that individual ants would not be able to on their own (cf. chapter 1).

specialisations, but he does deny that the brain is rigidly, functionally structured:¹⁷⁷

Lesions of the primary projection areas will of course affect the related functions, not because these areas control the functions, but rather because the relevant information is prevented from entering and leaving the system. The primary areas are not the place where things happen but rather, the distribution points from where information is disseminated throughout the brain (1989:69).

Cilliers substantiates this view through highlighting findings by Mountcastle in the 1960's (1989:70-76): cells in the cortex are organised into groups of cells that are heavily interconnected vertically and sparsely interconnected horizontally. Structurally such a mini-column consists of target neurones that receive impulses from the subcortical regions, target neurones that receive impulses from other areas in the cortex, local circuit neurones that integrate inputs, and output neurones that connect different columns and project impulses back to subcortical regions (73). Groups of columns work together to form a module. It is important to stipulate that modules do not work in isolation. Modules work together in large systems, with projections that form massive loops to the subcortical areas and back (73). Mountcastle found that the basic structure of these columns are repeated throughout the cortex, and he concludes that the brain functions as a distributed system, where information can follow many different routes, and where information can be cycled in loops. The dominance of one path or another in the flow of information becomes a dynamic property of the system and changes as properties of the system change. As a result, Mountcastle finds that distributed systems display redundant possible loci of command, and that the

¹⁷⁷ This study does not allow space for discussing some of the interesting experiments with regard to the specialisation of areas in the brain. The case of "split-brain" subjects is particularly interesting in the context of this study. Cf. (Trevarthen 1998b:740-746) for a concise discussion and recommended further reading.

command function changes between different loci within the system.¹⁷⁸ This characteristic allows for the adaptive capacity displayed by the neuronal apparatus, for example the possibility to regain and improve competence in function after brain lesions (74).

Cilliers (1989:74-75) briefly discusses the similarities between Mountcastle's contentions and Pribram's holographic model of the brain (mind).¹⁷⁹ Because, in a hologram, there is no one-to-one correspondence between spatial orientation of a piece of information on the original, information is distributed all over the hologram. Each part of the hologram contains information about the whole of the original object. The implication of this is twofold: on the one hand a small segment of a hologram can be used to recreate the original image (with the loss of detail and definition accordingly). Secondly, when a piece of the hologram is damaged, there will not be a corresponding hole in the image that it creates. Damage to the hologram would cause the clarity of the image to deteriorate proportionately to the damage (1989:75). The implication of the holographic theory of the brain with regard to memory is that "the brain does not store information in terms of simple space-time or causal relations" (95). Memory, here, is the result of

¹⁷⁸ Cf. (Cilliers 1989:74-76) where he briefly discusses similarities between Mountcastle's contentions and Karl Pribram's Holographic descriptions of the brain.

¹⁷⁹ Cilliers gives the following, simplified, explanation of how holograms work:

If an object is illuminated by a coherent light (like a laser beam), and the reflection is passed through a lens system, an image can be recorded (at the appropriate place) that bears no resemblance to the original object. As a matter of fact, it looks merely like a collection of finely curved lines and ripples, called "interference fringes." This is what is called a hologram. A replica of the original object can be recreated – in three dimensional space – by illuminating the hologram with a light beam resembling that with which it was created in the first place. The hologram is a transformation of the visual characteristics of the object that can be reversed, that is, an image of the original can be regained by transforming the transformation. The hologram contains all the optic information present in the object, but transformed into another mathematical dimension by means of the lenses. Because the transformation can be described fully in purely mathematical terms, lenses are not a prerequisite any more. The original information can be transformed into a hologram by calculation only. The massive amount of information to be processed necessitates the use of computers and results in what is known as computer generated holograms. This is mentioned here for the following reasons: Because the process can be done by means of a mathematical transformation only, the original information need not be visual information, but can be sound, or pressure or heat, or any kind of information that can be presented in the correct format to the mathematical transformer, and

external patterns being transformed into wave-patterns or ("coded" information), and the interaction of these wave-patterns resulting in interference patterns.

All of this leads Cilliers to discredit the idea of strict localisation of brain functions. A holographic theory on the other hand, would be:

...capable of explaining facts about the brain that otherwise seem baffling. For example, how is it possible that large parts of the brain can be destroyed, yet it does not seem to affect the performance concerning specific instances seriously. Lesions of the visual cortex for example, may impair certain visual functions, like creating blind spots, but will not make any specific object unrecognisable. This can be explained if information and images are stored and processed in the brain, not by means of pictures, or by some direct representation with a one-to-one correspondence, but by a holographic-like transformation that would involve the whole brain.

Cilliers is careful to emphasise that in lieu of rigid structure, he does not imagine the brain to be an amorphous structure. He insists that the brain has an architecture and that structure plays an important role in its functioning. His point, however, is that given structures in the brain do not work on given problems in isolation, and that in all probability there is a contingency, rather than a necessity to some of this structure (88-89).¹⁸⁰

although the result will not be a hologram in the strict sense, it will have the properties of a hologram (1989:74).

3. Where to from here?

Taking the preceding facts into account, it seems that Cilliers has a strong case for viewing the brain as a distributed system.¹⁸¹ In highly distributed systems, information is dispersed over the whole of the system, as information is highly dispersed throughout the structure of the brain (76). This, coupled with the argument discussed earlier that neuronal functioning is determined by the self-organising relationships between the components (neurones and their interconnections) in a structure (the nervous system, influenced by both genetic predispositions and by stimuli), serves as a departure point for Cilliers's theory of the brain as a distributed, open (i.e. complex) system.¹⁸² Such a distributed system's complex functioning provides the physiological underpinning for certain emergent properties that are not reducible to the components of the system as such, but results from the *interactions* between the components of the system. In such a scenario it becomes possible to propose a theory of the higher functions of the brain, without having to incorporate fantastical entities or metaphysical assumptions or having to be crudely reductionistic and try to localise these functions in some specific area of the brain.¹⁸³ Cilliers's attempt to give a materialistic account of the higher functions of the brain leads him to conclude that: "Consciousness is an emergent property of the brain" (Cilliers: private communication).¹⁸⁴

¹⁸⁰ Cf. complimentary findings made in studies on both "split-brain" and "normal" subjects (Trevarthen 1998b:740-746).

¹⁸¹ Also see Cilliers 1998:70-76.

¹⁸² The brain shares many of its structural characteristics with the post-structural take on the structure of language, especially that that stems from Saussure's linguistic models and their extension by the likes of Jacques Derrida. Seeing that both language and the brain can be described as open systems of differences, in other words, systems that consist of components that have no individual identity and derive function or meaning from their relationships with the other components of the system, Cilliers believes that models of language can inform models of the brain (1989:9; 86-98; 152-171; 189-191). See Freud and the Scene of Writing (Derrida 1978:196), where Derrida uses a post structural description of language to interpret Freud's model of the mental apparatus (see also Cilliers 1989:172-177; 1998:37-47).

¹⁸³ Cf. Gregory (1998b:217-218) for an short, but edifying discussion of the role played by emergence and reduction in explanations.

¹⁸⁴ Roger Sperry, famous for his experimental studies on how brain circuits are formed, and his research on "split-brain" patients came to a similar conclusion. His research

If one adheres to this conclusion, it is a very small step to take to conclude that what we call “self”, in its turn, is an emergent property of a complex system. All that remains to be done is to consider what this system might be, and to convincingly argue the case that self, as such, does indeed lend itself to a similar, materialistic, description.

In a sense the step has already been taken in the previous chapter with Dennett’s materialistic theory of the self. What remains to be done in this chapter is to present an argument establishing the self as the emergent property of a complex system, and indicating why and how such a theory of self would differ from that of Dennett. The final chapter will then be devoted to bringing to light possible implications our new, revised theory of self might have.

4. Dennett’s material self

Dennett’s recognises that his theory of consciousness holds important implications for the self, which leads him to consider a revised (with respect to the Cartesian position, especially) theory of self:

Without the Cartesian Theatre or Central Headquarters there is no single, definitive stream of consciousness. Instead, Dennett envisions multiple channels within the “virtual machine” in the brain, which create various narrative drafts that may or may not play functional roles in the activities of the brain (254). It seems to him to be a remarkable fact of nature that each member of the primate *Homo sapiens* constructs a self (416). Dennett refers to the self as a web of words and deeds and likens the inclination to construct a self on the part of human beings to the inclinations of spiders to spin webs or bowerbirds to construct elaborate bowers – a self is crucial to our success as human beings (416). Dennett also uses the image of the beaver constructing a protective fortress through actively gathering the necessarily

convinced him that most cerebral functions are genetically determined by a kind of chemical or physiochemical encoding of neuronal pathways and connections. Sperry’s research took a philosophical turn and he proposed a “monist theory of mind in which consciousness is conceived as an emergent, self-regulatory property of neural networks, which enables them to achieve certain built-in goals” (Trevvarthen 1998:114-117).

material and sharing the labour among individuals of the species. In like manner, humans appropriate all manner of “found-objects,” with which, among other things, to construct their own protective web of self. Dennett insists that the self is as much a biological product as any of the other constructions belonging to animals (*ibid.*). Human beings are constantly engaged in presenting themselves to other human beings and to themselves, through the ubiquitous medium of language. We tell stories of who we are, where we come from and where we are going – the result is what Dennett calls this our “narrative selfhood” (418).

Dennett’s conception of self is above all an abstraction (368), an effect rather than an internal boss, built up out of a myriad of attributions and interpretations composed into a narrative. Dennett calls such a narrative the “centre of gravity” of the agent. The active body is then in a position to include a mental model of itself in its representation of its environment, which is essential for long-term planning. By having a model of itself the body can also keep track of its internal states, decisions, tendencies, etc. (427-428).

Dennett is very taken with Dawkins’s theory of memes¹⁸⁵ and presents these narrative selfhoods as being spun by selfish memes in a quest for evolutionary advantage, and ultimately: survival. Such a stance leads him to declare: “Our tales are spun, but for the most part we don’t spin them; they spin us” (418). Although Dennett’s fundamental sentiment in making such an assertion may be correct, he runs the risk of representing memes as if they were executive agents with a teleological agenda and the ability to implement such an agenda. Presenting some kind of hominuculi that act as miniature agents within the mind is exactly what we are trying to avoid with this study. It is to this extent that complexity theory could prove a useful tool to develop a theory of self that is complimentary with that of Dennett, but hopefully avoids delegating “agency” to any part of the system that makes up the self. The argument would be that within a system of differences, order and structure would develop that give rise to the self, purely as a result of the characteristics

¹⁸⁵ Memes will be discussed in greater detail in the final chapter. Briefly, memes are the cultural equivalent of genes in biology and follow the same basic evolutionary principles.

and capacities of the system. At no point will it be necessary to introduce either a further entity with an agenda of its own, or some kind of centre of executive control that runs the operations of the brain in a top-down manner. What remains to be done in the present chapter is to explore some of the advantages to be had from modelling the self on the principles of complexity theory.

5. The self as a complex system

So far in this chapter we have explored a theory that roots one of the functions of the brain that is usually ascribed to mind, in the material aspects of brain structure and operation. The conclusion was that consciousness is an emergent property of the physiological workings of the brain as a complex, distributed system. We also explored the implication of Dennett's attempt at undermining mind-body dualism. Dennett's conclusions on consciousness, i.e. as the effect of multiple narrative drafts within the "virtual machine" of the brain, leads him to consider the implications that such a revised theory of consciousness would have on conceptions of self. The self ceases to be a fixed, coherent entity – the sentient headquarters that initiates and supervises brain processes. The self becomes an abstraction, a narrative construct that a human being creates in order to orientate him- or herself in the world. Essentially Dennett's conception of self is very convincing. But, by combining his material self with Cilliers's complex model of consciousness, a more plausible and less problematic theory of self seems possible. It could prove to be an especially useful way to comprehensively incorporate the environment into our conceptions of the construction and maintenance of self.

The proposed conception of the self as a material construct certainly coincides with all of the criteria necessary to be considered a complex, distributed system. In chapter 1 complex systems were discussed in some detail, and the following characteristics were identified as typical of complex systems:

Complex systems are open systems that are made up of many different components that are constituted by their interaction with one another

and with their environment. Such a system possesses emergent properties – in other words, the individual parts of the system do not aggregate into the whole system. The characteristics of the system are the result of the non-linear interactions between the components of the system amongst themselves and with their environment. A complex system cannot exist in isolation, it feeds off information and energy that flows into it from the external world, and on its part contributes to the external world. The boundaries between the system and its environment become difficult to determine because of this dynamic interaction. Due to the dynamic character of the system, only certain aspects of the system can be analysed at a time. Hence, in order to facilitate analysis an artificial frame needs to be imposed on the system, which creates the possibility that any analysis could cause significant distortions in the system and that such a possibility should be taken into account in the final analysis.

5.1 Cartesian conceptions of the self disqualified.

In light of the discussion up to this point it seems fair to assert that Cartesian conceptions of the mind (and by implication the self) do not by any stretch of the imagination present the mind as an open, interactive system. The difficulties in coherently delineating precisely what the differences are between the mind, self, ego, consciousness, etc. have already been discussed. For the purposes of this discussion it is possible to use the terms mind and self particularly, relatively interchangeably. Both concepts seem to be subsidiaries of the same overall concept of the characteristics associated with the higher functions, presumably unique to human beings, namely: consciousness of ourselves as agents within a greater world and with the ability to actively orient ourselves with regard to that awareness. To this end we need to have a “mental” pictures of the entity that is “us” which we can reflect upon and factor into our calculations regarding willed action.

The Cartesian picture of the mind, then, is a far cry from what we have described as a complex, distributed system: Descartes’ disembodied cogito needs to withdraw from the senses, seeing that Descartes readily accepts the

possibility that our senses, in cahoots with a malignant demon, may comprehensively deceive us in all that we perceive. To avoid such deception Descartes prefers to attend to those perceptions that have “sprung up of themselves” in his mind (1978:86-87). As we have seen, Descartes concludes that his cogito, his essential *self* is limited to being a “thinking thing” – a thing that doubts, understands [conceives], affirms, denies, wills, refuses, imagines, and perceives, in complete isolation from the body and world in which it resides. Descartes can *distinctly perceive* mind and body as two separate things and he takes this to be sufficient evidence that they are substantially different from one another and he knows with certainty that he exists as a mind. He has no reason to believe that his essence necessitates anything beyond being a thinking thing. He even concludes that his mind can exist without his body (132-133). Descartes assumption that he can attain universally valid knowledge makes it unproblematic for him to generalise attributes found within himself and apply them to all people.

We have also seen that despite contemporary and subsequent criticism, Cartesian dualism still pervasive in thought on the mind/self. Hence, although Hume's project was to establish how we come to have knowledge, based on observation and reasonable inference, rather than through reason alone, his speculations are not all that different from Cartesian dualism. Hume expels innate ideas, and in Hume's world perceptions, including those of the imagination, are limited to previous sensory experience. One cannot conjure up an original idea without any mental contents, and mental contents are all subsequent to experience. As discussed in Chapter 3, this assumption causes Hume to run into difficulty when it comes to conceiving of abstract concepts, which leads him to conclude that the self is merely the product of habit and custom. It suits us to conceive of a coherent self, because we have been accustomed to do so; it simplifies our comprehension of impressions. But, the self, as such, does not exist. As we have seen, Hume does not really escape Cartesian dualism, in fact, that was not his aim. As already discussed the dispute between the rationalist and empiricist positions basically boils down to whether reason or experience ultimately justifies belief. The dispute is over how knowledge is acquired and not about the merits of the underlying mind-

body distinction. Rationalist theory holds that the mind has innate ideas that form the basis on which reason can form justifiable beliefs about the world. Empiricism disputes the possibility that innate ideas exist and explores the limitations of reason in forming justifiable ideas about the world.

Hume does advocate the importance of individual character – a good upbringing, cultivation of the virtues, a respect for traditions – and believes that our sentiments and our natural common sense, cultivated through our social traditions, have power and virtue to address the limits of reason. In this sense Hume's theory moves more toward a conception of the "self" as a constructed system, and even an open system in interaction with its environment.

Kant continues the debate and finds unity in the *I think*, that must accompany all representations in the mind. He bases this assertion on the fact that if the *I think* did not accompany all representations, it would be possible for him to have something represented in him, without him thinking it, which he deems to be either impossible or at the very least, nothing in relation to him. With such a formulation, Kant also seems to subscribe to the Cartesian mind. The mind, the *I* that thinks, must be the author of representations (This image puts one in mind of Dennett's image of the Cartesian theatre). Non-"authorised" presentations are either illogical or negligible. Although Kant sets about establishing the conditions necessary for cognition, which need *not* rely on habit and custom and which can be justified objectively, one of his invaluable contributions to Western thought would be his recognition of our own input into our perception of the world. His conclusion would be that a self-consciousness must *necessarily* accompany every mental act, which follows that the self must exist. Essentially though, Kant's self is an operation of understanding, a permanent substratum and not a dynamic entity, subject to development. The open, dynamic and decentralised system that we are looking for still seems a long way off.

As discussed in Chapter 3 Freud provides us with an example of a decisive break with the Cartesian mind. He also makes an invaluable step decisively towards a conception of mind/self as a complex, open, distributed system (although, of course, he does not use this terminology). Freud's model

of the mental apparatus in the Project is an example of a classic complex system.¹⁸⁶

We have seen that Freud suspected that, far from being, clear, “certain, and indubitable” conscious, and by implication rational, thought processes were the result of intricacies of the body’s physiological functioning. Consequently he set about modelling the mental apparatus, and providing a physiological description of “mental” phenomena. He knew that the nervous system consisted of distinct and similarly constructed neurones, which have contact with one another through the medium of a foreign substance. He also knew that certain lines of conduction were laid down between neurones and that they receive and give off excitations.

Freud realised that in order to account for our ability to learn through experience, somehow memories of experiences need to be stored and available for a kind of mental cross-referencing in future situations. He speculated, and as we have seen he was quite accurate, that certain pathways in the nervous system are strengthened through use and thus become more likely to be “fixed” within the neuronal structure. Memory (the capacity to be permanently altered through experiences) becomes the basic property of neurones and their interconnections. It becomes the basic component of the nervous system and is prior to consciousness and cognition. Note that Freud’s description of how memory functions, is (unbeknownst to him, of course) itself a description of a complex system. The following quotation serves as a good example: “Memory does not lie in the facilitated pathways themselves, but in the relationship between them, and this relationship is one of differences” (Cilliers 1998:112).

From here it is just a small step for Freud to launch us into the realm of the unconscious. The procedure of forming memory traces is entirely unconscious. Whereas “every psychological theory ... should explain to us what we are aware of, in the most puzzling fashion, through our “consciousness” and, since this

¹⁸⁶ See Cilliers and Gouws (2001:237-256) for an extended argument supporting a similar conclusion.

consciousness knows nothing of what we have so far been assuming – quantities and neurones – it should explain this lack of knowledge to us as well” (Freud 1950 [1895]:307-308).

It becomes clear to Freud that if one is prepared to find that some assumptions are not confirmed through consciousness, and that consciousness does not provide trustworthy knowledge of neuronal processes, even though they seem to be the origin of consciousness, that neuronal processes are in fact *unconscious* processes that can be approached from a scientific angle, “and inferred like other natural things” (1950 [1895]:307-308). And, if neuronal processes themselves are on the whole unconscious processes, the likelihood arises that much of our thought processes are in fact unconscious. Memories are in themselves unconscious and, although having the capacity for becoming conscious, can be as influential while being in an unconscious state as in a conscious state, which is of course one of the premises that psychoanalysis is based upon.

As we have seen, Freud recognises the immense implication that his theories on the unconscious would have on theories of mind and insists that the only way in which to form “a correct view of the origin” of mental phenomena, is to not place too great a stake in the property of “becoming conscious” (612). The advent of the unconscious expanded the possible spheres of mental life exponentially by doing away with the assumption that the psychical is equivalent to the conscious.

Freud's consciousness is partly subjective and partly the result of external conditions. Consciousness becomes, not the property of a specific system, but the result of the *interaction* between perception and memory. Having established the basic mechanism for generating consciousness, Freud proceeds to explain a number of the “higher” psychological functions in terms of the existence of the unconscious.

Freud's realisation provides us with the necessary conceptual tools that seem to be wanting when studying phenomena like consciousness and the self through introspection and speculation. His methods of studying dreams,

slips of the tongue etc., and the idea that in some instances, an external observer, like a psychologist, might have more insight into a person's motivations than the person him/herself opened a wealth of opportunities to rethink many received wisdoms about the mind. It becomes possible in this study to speculate upon some of the possible internal processes that produce the self, and to assert that these processes are unquestionably similar to the processes typical to complex systems. Unconscious processes, consisting in part of memories (external influences), which themselves are the result of systemic differences, and processes that may under certain circumstances become conscious, but need not have an enormous influence on mental processes, lend themselves to being modelled as complex systems. Furthermore, Freud also provides us with some clues as to the greater system in which the self is likely to emerge.

5.2 Freud lays the foundation

We have seen that Freud speculates that the mental apparatus is set in motion because the accumulation of quantity ($Q\acute{n}$), as stimulus that arises as a result of the needs of the body, will create an urgency in ψ -system to discharge this energy. The only way for the pressure to be relieved is to effect an external change to get rid of the stimulus that causes the release of $Q\acute{n}$ in the interior of the body. The fact that the infant human organism is incapable of bringing about the necessary action, and needs to take recourse to external help, is a very important aspect of Freud's theory. In order to survive, and in order to affect the necessary changes in its mental apparatus, the infant is initially completely dependent on a caretaker. The child's internal states need to be communicated to the caretaker, which is accomplished "by discharge along the path of internal change" (1950 [1895]:318). The caretaker is, from the first, the most important environmental influence on mental development. We have seen that Freud believes that this initial helplessness of human beings is the *primal* source of our moral motives. Freud does not elaborate on the relationship between the infant and the caretaker, but the meaning he attaches to this initial and inevitable encounter seems to indicate that he

believes social interaction to be a *necessary* precondition for the development of consciousness.

We have seen that Cilliers (1989:118) explicitly draws the conclusion from Freud's proposed process that the body, as such, and its needs are necessary precursors to consciousness. Bodily needs, or endogenous stimuli, become one of the primary movers of the physical system.

Clearly Freud's mental apparatus is an instance of an open system – a system in continual interaction with its environment. In fact, as stated in chapter 1, open systems exist *by virtue* of their interaction with their environment. Such systems exchange matter and energy (and information) with the environment, to such an extent that it becomes almost impossible to delineate between “system” and “environment”.

According to Freud consciousness, as an emergent property of the complex system of the human organism and its environment, serves an important function in this context. Consciousness becomes a kind of reality detector. As discussed, Freud speculates that initially the mental apparatus might try to rid itself of accumulated stimuli when the body's needs are not met by means of wishful hallucinations (see chapter 3). It seems logical that the psychical apparatus needs to develop a conception of what really is the case in the external world, rather than forming a presentation that is merely agreeable, in order to have the needs of the body met. With hallucination as a means of dealing with accumulating stimuli is no longer a viable option for the mental apparatus there arises the necessity for a new mental function. This mental function is introduced as Freud's famous *reality principle*.

In the reality-detection game *consciousness* becomes a key player. Consciousness is augmented with the function of *attention*, and attention, on its part, depends on the existence of the faculty of *memory*. Consciousness would hopefully then be in a position to make impartial judgements on the truth or falsity of given ideas, by comparing the ideas to memory-traces of previously encountered reality and deciding whether the ideas are in agreement with reality or not. After this judgement motor discharge can be allotted with the function of action, in other words with the task of altering

reality and thereby relieving the body of excess stimuli. As we have already seen in our discussion of the *Project*, *thinking* developed from the presentation of ideas, and arose out of the need to allow the mental apparatus to tolerate the increased strain put on it, in the event that discharge is postponed. Here we have an explanation for the basic mechanism of adaptation and learning. And, as discussed in chapter 1, the possibilities for emergence are compounded when elements of the system allow for some capacity for adaptation and learning.

Here, already, we have a theory of a complex system, where certain properties of the system, like consciousness, emerge from the interaction between the components of the system (and the environment). These processes are not mere epi-phenomena to the system. They feed back into the system and in themselves become a significant part of the functioning of the system. Freud's description accords very well with the broad definition of emergence given in chapter 1 as an instance where: "a small number of rules can generate a system of surprising complexity" (Cilliers1998:3). We know that emergence belongs to the structural aspect of the system and that the system does not need to have certain kinds of constituents or mechanics to have emergent properties. We know that many very different kinds of systems can exhibit emergent properties and that such systems are not limited to physical systems and can include social, cultural and biological systems as well.

Freud describes the self in the following way: "What we describe as our 'character' is based on the memory-traces of our impressions which have had the greatest effect on us – those of our earliest youth – are precisely the ones which scarcely ever become conscious" (540). Not only does Freud raise the possibility that much of what we regard as the self is, in fact, unconscious, and hence not necessarily "known" to us, but he lays the foundation for modelling the self as a complex system. The only shortcoming in Freud's description seems to be that he envisions too narrow a sphere of influences that can possibly contribute to make up "character". This study would contend that there are many more factors than the memory-traces of the experiences of our youth contribute to "what we describe as our "character".

5.3 The material basis of the self revisited

Earlier in this chapter we discussed some of the similarities between current neurological theory and the Freudian model. Especially in terms of learning and adaptation we have seen that great emphasis is placed on the role of environmental influences in forming the structure of the brain and its processes. Trevarthen (1998:102), for example, places much emphasis on the interaction between genetic predisposition and environmental stimuli in the development of the brain, including the development of the “higher psychological processes.” We have also quoted him as declaring that after contemporary research done on this subject, the way in which we think about consciousness and human understanding can never be the same (i.e. Cartesian). Trevarthen attributes some of the diversity of human minds to genetic pre-programming but he insists that *the same* processes will be influenced by stimuli from both the intrauterine and the external environments as well (*ibid.*)

The discussion on current neurological theory also touched upon the fact that the organism needs to be able to learn from experience. In order for the human organism to have *learned* from experience information acquired from experience needs to be incorporated into, and has to influence, the workings of the brain. As with Freud, Young’s discussion on learning hinges around changes brought about in the mental apparatus when learning. Briefly, he argues that to be able to learn, we need physical records in the brain – changes need to be brought about in the structure of the brain, i.e. memories. The problem of memory is finding the mechanism that establishes this change. We have also seen that Young’s (1998:455) physiological account of the mechanisms of memory begins with hereditary genetic dispositions. The initial basis of memory in the nervous system is provided by “genetic memory”, which establishes “pre-wired” neuronal pathways of the foetus during gestation. After birth this original multiplicity of possible actions is then “fine-tuned” through interactions with the environment: redundant neuronal pathways are pruned away, while others are strengthened. An important aspect of Freud’s theory on memory is that he searches for a metaphor that

illustrates that memories are never completely lost, nor are they saved as complete icons in some filing-system of the brain. The metaphor that he settles on is, of course, that of the Mystic Writing-Pad. We found a relatively similar account of memory in our discussion of the holographic model of memory. Here memory is the result of external patterns being transformed into wave-patterns or ("coded" information), and the interaction of these wave-patterns resulting in interference patterns. What both of these descriptions have in common is a view of memory as non-iconic and distributed in the brain.

As with Freud there is much emphasis on the role of the environment (which includes the caretaker or caretakers) on brain development after birth, in the neurological approach. After birth, brain development depends on how stimuli from the outside world interact with the inner states of the brain. To quote Trevarthen's evocative statement once again: "Growing human brains require cultivation by intimate communication with older human brains" (108). So much so that it seems that communication with older human brains is at least partly, if not completely, necessary to the development of consciousness.

When all of these ingredients are present (i.e. effective genetic pre-programming, effective caretaking, learning and adaptation) it seems that the system is in place for a conscious individual, with a sense of self to emerge. We have seen that, far from being amorphous or chaotic, complex systems have some structure, some steady-state where they are more or less coherent and can do work. From the patterns of interaction of a system in flux emerges a pattern (or perhaps patterns), which is relatively constant in its composition, with a more or less coherent structure.

We have seen that part and parcel of the dynamic nature of complex systems is their remarkable tendency to display *organisation*. The kind of organisation typical to complex systems is ideal for explaining the organisation that we can attribute to the self. Random patterns resulting from the interaction between components in a system do not make for emergent properties. Emergent properties are *ordered and recurring* patterns that come about through some kind of organisation among the components of that

particular system. We have also seen that the most probable state of distribution within a system is that of complete disorder, or maximum entropy, and that work needs to be done on the system to maintain structure within the system.

Evidently paradoxically, the patterns that are generated in complex systems are not the convolutions of random patterns. The resultant systems exhibit, as they must to be labelled a “system”, recognisable structures. These recognisable structures are dynamic – they change over time. In chapter 2 we saw that while the rules that generate the system stay invariant, the things that they govern are in flux – such is the nature of a *complex* system. In other words emergent phenomena are typically persistent patterns within a system with changing components. Relatively consistent rules (or behavioural laws) generate complexity and the flux of patterns that follow lead to “perpetual novelty” and emergence.

This apparent paradoxical aspect of complex systems serves to emphasise the importance of context-sensitive constraints on self-organisation in a complex system. An important effect of context-sensitive constraints is that they regulate the flow of energy, matter and information between the system and its environment. The organisation of the self-organised system determines the stimuli to which it will respond (as Freud’s model in the Project suggests with regard to the mental apparatus). The elements in a complex system are interdependent, which means that the behavioural variability they might have had as independent elements is constrained – this aspect could be very useful in a theory of the self. Context-sensitive constraints enable the system to preserve its organisation and its identity as a whole. Instead of some governing component determining and regulating the behaviour of the system, the relational whole of the system governs the behaviour of the system. The self as an emergent property of the genetic mental apparatus and external influences could well be said to be the result of self-organisation within this complex system. This assertion ties in very well with Dennett’s discussion on both consciousness and the self. And perhaps Dennett can also provide us with a clue as to how work is “done” on the system and how recognisable patterns (order) usually stay constant

enough for us to recognise a self as the “same” self at different instances and in different contexts.

5.4 Dennett’s self revisited

In chapter 3 we used Dennett’s materialism to underpin our efforts to demystify the self, and to present it as an aspect of consciousness, also emergent from the complexity of the brain’s structure and functioning. He strongly criticises that *Cartesian Theatre* model of consciousness, which poses some kind of homunculus as the executive commander of the mental system. As an alternative he suggests his “Multiple Drafts” model. According to the Multiple Drafts model all information entering the nervous system is under continuous “editorial revision”. Dennett insists that it is misleading to ask the question when perceptions become conscious. The information content gleaned from sensory inputs is distributed throughout different systems of the brain. This distributed content becomes something like a narrative stream – a multiplicity, subject to continual editing by many processes distributed in the brain. Hence there are multiple “drafts” of narrative fragments, based on sensory experience, at various stages of “editing” in various places in the brain. There is no single, final narrative, which is delivered to consciousness and can be considered to be the actual stream of consciousness of the subject. There is no point in the brain where it all comes together. The brain is the “headquarters” of the perceived observer, but there is no other, deeper, headquarters in the brain, where consciousness is seated.

Dennett’s new model of the conscious processes within the brain allows him to rethink and recast the self, which had mostly been conjectured upon in context of the assumption of the Cartesian Theatre as model for the “mental” aspects of the brain. We arrive at his conclusions on the self via a somewhat circuitous route:

The Multiple Drafts model lends itself to a theory of the evolution of consciousness. Far from being a metaphysical, non-bodily phenomenon consciousness becomes a biological effect that developed with a species,

presumably in accordance with constraints and possibilities imposed by the environment and genetic adaptations on that species.

We saw that Dennett speculates that the need for self-preservation and control through the ability to track and anticipate, were evolutionary preconditions that gave rise to the nervous system in successive guises. The ability to model the world, to learn from experience in order to adapt or anticipate future occurrences seems to give an organism a distinctive biological edge over species that do not have this capacity. Stating the matter very simplistically, organisms that have systems proficient in information gathering and geared towards information that is beneficial to the organism, are likely to prosper. Eventually it would seem that these information-gathering systems have become part of the innate design of the nervous system. Information gathering and assessment need not necessarily be conscious states.

Specifically with regard to the self, Dennett's discussion holds yet another point of interest for us. He notes that the development of nervous systems that have an element of plasticity and hence have the ability to learn in the course of their lifetime seems to provide other "mediums" in which evolution of the nervous system can occur. Such a learning mechanism would operate along the same lines as "natural" evolution, in other words, a process of evolution through selection. Dennett refers to this process as *post-natal design fixing*. We have already discussed how the plasticity of the brain allows it to reorganise itself in some ways, even if these ways are constrained (!), and so adapt to its environment.

The process of the evolution of the nervous system, which has hitherto been seen as been driven by natural selection and genetic mutation, might be said to evolve in other ways as well. And Dennett goes on to suggest that what makes the human nervous system unique is that it has developed the ability to conceive of a quite sophisticated model of itself.

Dennett's speculation is strengthened in the light of the assertion on the part of Nathan that the cerebral cortex is a late evolutionary development, and especially well-developed in man. The cerebral hemispheres are the part of the brain concerned with the activities usually categorised as mental:

problem-solving, remembering, planning, imagining, making judgements, forming opinions, etc. The cerebral hemispheres are also the regions of the brain that can differ overtly from person to person (Nathan 1998:531).

According to Dennett then, a plastic, adaptable brain (the cortex) is the first “new” medium in which the evolutionary process with regard to nervous systems can be speeded up. We have seen that Dennett attributes the radical transformation of human society in the last 10,000 years to the development of new ways in harnessing mental capabilities.

The second new type of evolution that Dennett discusses, is *cultural* evolution. This evolutionary medium is the product of both the plasticity of the brain, which makes learning possible, and human beings’ communicative capabilities. Through cultural transmission we transfer behavioural patterns to developing young minds, through some or other form of language. In fact, we quoted Dennett as stating that through such programme or software-sharing, culture develops into “a repository and transmission medium for innovations” (199). And so, as with Freud, Dennett places great emphasis on the importance that culture has for the existence and development of the mental structure. Finally, the cultural equivalent of genes, memes, in themselves undergo a process of evolution. They implement themselves in the human nervous system through cultural transmission and so also affect changes within this system. All of Dennett’s proposed evolutionary processes to do with the mental apparatus are important to our theory of the self.

Dennett’s Multiple Draft model of the mental apparatus allows for the self to be a sort of cumulative narrative draft, composed of myriad bits of narrative fragments formed in the brain. Dennett does, to a certain extent, attempt to account for the formation of such a self-like narrative draft. As already noted he suggests that human beings are biologically given to a process of self-design, which is as innate to their nature as it is to a beaver to build a dam or a weaver to construct a nest. One of the first steps in this process of self-design, after birth, is to acquire language. Dennett goes so far as to suggest that prior to language the self, in any meaningful sense of the word, does not exist, while the capacity for designing and developing a self does exist. This is reminiscent of Freud’s and Trevarthen’s emphasis on the

importance of language and communication in general for developing consciousness. By incorporating the three processes of the evolution of consciousness, and especially the role that Dawkin's memes may play in the process Dennett has provided us with the last of the conceptual tools necessary to construct a complex, materialistic model of the self. Dennett's self is alive, were "alive" can be understood in the sense of being a system that can sustain itself, evolve and reproduce.

5.5 The material and complex self

In terms of the criteria for identifying a system as a complex system we now have all the ingredients necessary to propose that the self is an emergent property of a complex system. We have the many components that are in interaction – genetic predisposition to construct a self, the mental apparatus itself, structured to learn and adapt, memories of experiences, other people, language and memes. From our discussion up to this point it becomes clear that these components are in interaction – the environment contributes to and influences the development of the mental apparatus, etc. The individual parts do not aggregate into the whole system – all of these parts can only lead to the emergent phenomenon of the self when they interact effectively, as it were. In the vernacular of complexity theory, the law of superposition does not hold. In terms of Freudian theory it is easy to propose that causes and effects within the mental apparatus are highly disproportionate, unstable and unpredictable. In other words, slight perturbations propagate through and affect the entire system, which can result in a system with behaviour that can be what Auyang calls "multifarious, unstable and surprising" (1998:183).

As we have seen in chapter 1 such non-linearity is a necessary, but not sufficient precondition for complexity. Freud himself places much emphasis on the impact that experiences in infancy, both conscious and unconscious (especially of a sexual nature) might have on the grown person. Such effects would of course vary from person to person, and vary in the same person at different times, based on all manner of contingencies and systemic variations. One does not need to be quite as fatalistic or as focussed on sexual

experience as Freud seems to be, to appreciate that the same basic principle can hold for most human experiences throughout their lifetimes.

The mental apparatus consists of various structures at various scales within the brain. This makes the concept of coarse-graining particularly applicable to the self-system – the distance we take from the system in analysing it influences appearance of complexity/complicatedness that we perceive the system to possess. The other side of this coin is that there are more possibilities to the development of the self than can ever be actualised. We have quoted Auyang as saying that the difference between the enormity of the possibilities in a complex system and the scarcity of actualisations underpins concepts like probability, contingency, temporal irreversibility and uncertainty (18) – all concepts that seem to be tailor-made to be applied to the self. It seems very easy to imagine that one's ownness, one's sense of self could be very different given the possibility that one had made different choices, or had different influences and exposure to different circumstances than what had actually been the case. All of these factors have to do with restrictions within the given system within which one constructs a self: physical ability, genetic-predisposition, material circumstances, geographical location, the list is endless. It is also quite unproblematic to state that many of the possibilities that have been actualised, much of what has become part of one's self are, for all intents and purposes, irreversible.

Given the enormity of the possibilities within the self as a complex system, it begins to seem surprising that we seem to possess something that seems like a coherent self, recognisable as belonging to the "same" person. As we have seen of emergent properties, they are the product of interactions between agents (nodes) within the system and dependent on context. In chapter 1 we quoted Holland as proposing that the context in which an emergent pattern arises determines its function (1998: 121-226). Could one count on the constraints provided by the context in which a self develops to provide the basis for a coherent self, recognisable as "the same" structure over time and through various circumstances and experiences? In fact, complexity theory seems to provide us with a very handy way of accounting

for such identity of self, while still allowing for profound changes to occur in the self, both diachronically and synchronically.

We have seen that open systems change continuously and in unpredictable ways. Components are added and subtracted and the relationships among various components change, which, in turn, influences the rest of the system. Open systems do not operate according to the dictates of a *telos*, changes are unpredictable, irregular, and contingent. The self does not possess some Aristotelian formal cause or essence, which dictates its development. We have also attributed the characteristic of decentralised decision making to the self. No specific component controls the system, and with changes that are the result of many factors, including changes in the environment and in other distant parts of the system, we end up with a mercurial and transient system.

As we have discussed, there are many advantages to not being directed by a single centre of control, but a system must, by definition, display coordinated and interactive behaviour. With the complexity sciences it becomes possible to explain how a complex system can attain such a high degree of order, without some external designing or directing agent. In chapter 1 we introduced the phenomenon of *self-organisation*. Self-organisation is a profoundly pragmatic, prosaic occurrence that has to do with the optimal functioning of a system. It will also help us to do away with the idea of self as essence or some such innate, non-physical entity.

In a self-organised system the individual components of the system react to information available to them locally, which translates into complex and organised behaviour on a systemic level. Any one node is not “aware” of the behaviour and structure of the entire system. The system itself relies on interactive behaviour and is subject to continuous change. Self-organising systems are also self-referential in that new components are “accepted” into the system by virtue of their ability to enhance the overall organisation of the system. We have seen that the system’s organisation makes for an internal selection process, established by the system itself, and operates to preserve and enhance the system. In other words, self-organising processes are primarily informational. The components that are imported into the system are

selected through the internal dynamics of the system, based on the system's requirements. Under pressure from its environment, the system seeks to enhance its cohesion and integration and functioning capabilities.

We have also discussed the concept of *self-organised criticality* – the tendency of large systems with many components to evolve to an unbalanced, yet structured state – a “critical” – state, where minor disturbances may lead to events of all sizes. Cilliers (1998) explains what takes place in self-organisation as the system organising itself to a critical point where single events will have the widest possible range of effects and where the system can attain optimum sensitivity to external inputs. These critical points are also called “attractors” – states that a system eventually settles into, and which are determined by the properties of the system.

The state-space of a system is a “structured collection of all possible momentary states of the individual” (Auyang 1998:215). A system's states have different characteristics at different times. Successive states – or a system's trajectory – taken as a whole, constitute a system's history or the process that it undergoes to reach a certain state at a certain time. We have seen that Auyang adds her voice to those of Cilliers and Dennett when she argues for the impossibility to encompass an entire system (including its history) in state space. And, she proposes the same solution: resorting to narrative explanations when describing the system.

We have seen that every possible state of the system is characterised by a unique point in the state-space. The unfolding of the dynamic system through time forms trajectories through the state-space. When a number of trajectories converge on a certain point in the state-space that point is called an *attractor* – a stable state of the system. A characteristic that could be of great relevance to our theory of the self is that a stable system has only a *few* strong attractors, whereas an unstable system would have no strong attractors and would jump around chaotically. A chaotic system has no structure and is useless, while a stable system with *too few* attractors is very rigid and cannot readily adapt to changing conditions. It seems that an effective system will balance itself at a critical point between rigid order and chaos. At this critical point single events in the system can have the widest

possible range of effects, without disrupting the system. The system is able to evolve both by accumulation of small changes and by dramatic changes, in which evolutionary novelties can emerge, triggered by random mutations or by changes in the environment.

A system with a few strong attractors will be at its most sensitive to external input, without being unstable. The system will also be able to change its state with the least amount of effort and the least amount of disruption. A system with too many attractors is unstable, and any slight perturbation might send it into another basin of attraction, thus changing its cycle and disrupting the system's "pattern". A system in which all its attractors are unstable would be a chaotic system, vulnerable to all manner of fluctuations and never able to repeat its cycles through the state-space and not able to retain its order.

It becomes very tempting at this point to veer off and speculate on the possibilities that the concept of attractor states could hold for the description of healthy vs. pathological states within different individuals. This would, however, overreach the boundaries of this paper and is therefore just held forth as a possible, and interesting, implication that such a conception of self might have. Instead we will content ourselves with the proposition that the self seems to be an ideal candidate for being characterised as an attractor state, or more likely a few attractor states, within a complex system. Even more specifically, the self would be a *strange attractor* state within a complex system.

We have seen that strange attractors are ordered patterns within a system that will still allow individual behaviour to fluctuate. So even if trajectories in a system are caught in an attractor basin, their behaviour is not so rigidly constrained so as to cause the individual trajectories to repeat a phase *exactly*. Consequently, even though the system is constrained in a state-cycle and does display order, individual trajectories are never exactly identical, but approximate. Even though the trajectories of a system in such a state may appear to be relatively random, such intricate behaviour patterns are in fact indicative of dynamic, complex and context-dependent organisation.

Seeing the self as an instance of a strange attractor that emerges within the complex system consisting of components such as the hard-wired structure of the brain, external influences in the form of physical stimuli and communications made by other people through language, could explain a seemingly paradoxical character of the self. On the one hand it seems that a single person has many “selves” depending on social context, such as a professional meeting versus meeting with friends in a pub, their state of health – any manner of factors. At the same time one needs to account for the fact that we consider a person to be the “same” person from infancy through death, barring extremely traumatic experience such as serious injury or extreme mental illness. Even when a person does an about-face in the manner of Scrooge, one would usually assume that person to be the same, only different in some specific aspects. Given the scope for strange-attractors to allow for deviations from “normal” trajectories, these attractors can still be recognised as the same basin of attraction, despite many differences. This conjecture of course may fly somewhat in the face of the Freudian assertion that experiences in infancy and childhood determine certain future characteristics. According to a model of the self as strange attractor, the self is open to influences, adaptations and changes until the system reaches a state of equilibrium, i.e. death.

The possibilities that are actualised, the attractors that the system settles into, are established partly by the environment (and the function of the system in that environment), and partly by the history of the system, i.e. states that have already been actualised. “History” here refers to already actualised states of the system and should not be understood as a chronological series of major events. The history of the system is contained in all the individual little interactions that take place all the time and are distributed all over the system as a whole. Future developments are of course constrained by already actualised events in the system. As a self-organised system the self derives its identity from its context – constituent elements become dependent on the behaviour of their neighbouring elements as well as what happened previously in the system.

It almost seems a truism to assert that the self forms within the constraints and influences that it encounters both within its environment, and from what it has already been, its history. But this is hardly obvious to a rationalist approach to the self, and even Freud's conception of the self is somewhat deterministic and static. When we model the self as a complex system it becomes possible to allow for both the innate aspects of the self – the “hardware” of the brain that ensures the ability to construct a self, presumably because it evolved as something that contributes to the survival of the human organism – and for the external, experiential factors, such as parentage, culture, education, geographical locality, or any external information that could have contributed to the formation of the self. Such a theory would also allow for, and be sensitive to, the extraordinarily complex interactions of all of these factors, and many others, that contribute to a person's identity. All of these factors and processes would of course be so intricate and convoluted that it seems unlikely that one will ever be able to account for every single aspect of a self's history or even of its present state.

With the impossibility of constructing a static model of this entire system of the self and all its possible states and configurations, it becomes necessary to determine the appropriate level of detail at which the system will be approached and the mechanisms of the system which are relevant to a particular study. We discussed the inevitability of *framing* and its implication that, not being able to reduce the behaviour of a complex system to a set of basic laws does not mean that it cannot be modelled or studied. We *can* reduce the behaviour of the whole to the *lawful* behaviour of the parts, if we allow for the inevitable distortions in the system that will arise from this practice.

When we practise framing our attention is attracted by a recurring pattern in a particular system and through a process of induction we construct a model of the selected phenomena. Only recurring patterns in a given system will be noticed and considered to be part of the mechanics of the system. Such a description of repeated elements will suggest rules or mechanisms according to which the system operates. It is the imposed frame that creates the safe space where there can be talk of truth, rationality, and

identity (!) To have a theory about the self, for example, is inevitably to draw a frame. As long as we remain well within the frame our work can continue. Ambiguities arise at the limits of the discourse or frame.

Perhaps a process similar to framing is what we use to identify someone else's self. We inevitably know the person within a specific context, and will base our conception of that person on the recurring patterns that we observe and also might exacerbate in accordance with our own experience.

Only certain aspects of the system of the self can be analysed at a time and analyses could cause distortions in the system. Both Freud and Dennett have it made clear that what is conscious, and self-like about human beings is the tip of the ice-berg with regard to the unconscious goings on, all the multiple narrative drafts within the brain. We have seen that Dennett proposes a method of *heterophenomenology* to extract bits of narrative fragments, or texts from the speaking subject. Such texts are then used by the theorist to generate a *theorist's fiction*: the "heterophenomenological" world of the given subject. This fiction is an account of all that the subject sincerely believes to exist in his/her conscious experience and is, Dennett insists, an *accurate* portrayal of what it is like to be that subject. He also insists that such a fiction is an adequate basis from which to explore the heterophenomenological world of the subject. This is in opposition to theorists like Searle and Nagel who assume the position that the self cannot really be a subject of scientific study, because it does not lend itself to objective study in that a theorist has to rely on the subjective account of the object of study itself. Their objection seems to be nothing more than the objection that a theorist is likely to cause a distortion in the system that he/she is analysing. Dennett refuses to concede that such a distortion is an insurmountable impediment; he proposes drawing up a frame – that of the heterophenomenological method – to allow for inevitable distortions. Far from being prohibitive, the subject's narratives enable a theorist to construct a surprisingly accurate, and we would argue adequate, account of what it is like to be that subject.

Both Dennett and Freud place much emphasis on the role that language, both spoken and written, plays in structuring consciousness. Humans make use of not only language and social interaction, but writing and

diagramming and other ways of storing information as well, for storing what one can refer to as a sort of cumulative, collective, human consciousness. Culture serves to augment memory, on the one hand simply because the demands on memory and pattern recognition are so vast that the brain is required to “off-load” some of its memories into buffers in the environment. On the other hand, as Freud notes, all people – normal and neurotic – have reason to distrust their memory, and can guarantee its authenticity by some kind of supplement (1925:227). Culture provides us with permanent memory-traces and we are fairly certain of these external “memory-traces” to remain relatively unaltered and undistorted, which is not always the case with actual memory.

The importance of other “mature human brains” as Trevarthen has it, and the cultural entities through which different brains communicate with one another will take up the bulk of the discussion in the final chapter. In his discussion on Dawkin’s memes Dennett has already pointed out the vital role that the culture in which they manifest themselves plays in the evolution of consciousness. In the final chapter we will argue, given our conception of the self as an emergent property in a complex system, of which culture (as environment) makes up a great part, we will discuss some of the implications that culture might have on the structure of the self.

Chapter Five

On the Subject of the Human Environment

We biologists have assimilated the idea of evolution so deeply that we tend to forget that that it is only one of many possible kinds of evolutions.

Dawkins (1976:208)

The claim to be a creator, a maker of things, passed from the painter to the engineer – leaving to the artist only the small consolation of being a maker of dreams.

Gombrich (1968:83)

Events in the past have to be interpreted in an imaginative way. Story-telling is the most appropriate way of doing this. Stories about the past enable us to create and share a common future. They contribute to the production and consumption of an informed culture for it is through the art of story-telling that a culture is enriched with intertextual significance

(Degenaar 1993: 54).

1. Introduction

In the preceding chapter we proposed that the self is an emergent property of a complex system that consists of relevant elements of the mental apparatus, and the environment in which it finds itself – broadly conceptualised as culture. We have seen that both Freud and contemporary neurology both place much emphasis on the influence of other brains on the developing human brain. Other people are not only important in the formative years, in fact, they can be argued to play a vital role in keeping a brain healthy and active through regular interaction throughout a person's life-span. Both Freud and Dennett have characterised the self as something akin to multiple narrative drafts within the brain. We have also seen that Dennett proposes a method of *heterophenomenology* to extract bits of narrative fragments, or *texts*, from the speaking subject. Those texts are then used by the theorist, and people in general, to generate a *theorist's fiction* – perhaps one could just

say a *fiction* – as to “what it is like” for that subject to be. Far from seeing such a mode of access to someone’s thought-processes as an impediment to, what Dennett would call, software-sharing between brains, Dennett sees his heterophenomenological method as a perfectly adequate means of understanding and communicating with “other minds”. This is in opposition to theorists like Nagel, for instance, who assume the position that the self cannot really be a subject of scientific study, because it does not lend itself to the objective study of objective facts.

Both Dennett and Freud place much emphasis on the role that language, both spoken and written, plays in structuring consciousness. Language in its various forms (culture?) serves as a medium through which knowledge is transmitted between brains. We have even seen that Dennett sees culture in itself as a relatively new evolutionary medium that contributes another dimension to the evolution of consciousness. Culture provides us with permanent memory-traces, which can remain relatively unaltered and undistorted and serve as entities through which different brains communicate with one another. Given our conception of the self as an emergent property in a complex system, of which culture (as environment) makes up a great part, we will discuss some of the implications that culture might have on the structure of the self, and some of the forms that culture may assume.

A discussion on self/consciousness seems, inevitably, to transform into a discussion on knowledge. Especially in a discussion such as this one, where we present the self as something that is *acquired* – through experience, through learning, through training, through acquiring culture, if you will. But why should this be? How do our views on knowledge and our views on selfhood/consciousness converge? The answer lies in what we believe knowledge to be – how it is acquired, but also how it is verified, justified, proven. A clear illustration some of the issues involved is a debate which centres around Thomas Nagel’s well-known thought experiment entitled: *What is it like to be a Bat?* Nagel’s reservations about our ability to know anything about consciousness, and Hofstadter’s dismissal of Nagel’s qualms highlight the way that ideas on consciousness/self and knowledge are interwoven.

2. Nagel's mysterianism and bats

Essentially Nagel (1982:391) is very sceptical about our ability to say anything meaningful about consciousness (and by implication the self) and he believes that current discussions on the subject get it “obviously wrong.” He is especially wary of attempts to describe consciousness in material terms. For him material or physical descriptions involve reduction. And, although reduction has proved to be successful in other attempts to explain the physical world, Nagel insists that it cannot be used to describe mental phenomena for exactly that reason, they are not *physical*. Nagel concludes that we need to develop an “objective phenomenology” if we hope to ever give a material account of mental events.

Nagel declares that most reductionist attempts to reduce mental phenomena to a variant of materialism fail, because they do not appreciate the distinctive difference between the mind-body problem and other problems that successfully lend themselves to reduction – that of consciousness. Nagel insists that mental events are subjective – inevitably connected to a particular point of view – and hence do not lend themselves to reduction. If material accounts of the mind hope to be successful, they need to account for this subjective character of consciousness. They need to light upon a way to describe subjective experience in objective terms. Why? Because legitimate knowledge needs to be objectively verifiable.

Nagel proceeds to discuss consciousness, even though he admits that: “it is difficult to say in general what provides evidence of it” (392). These qualms aside, Nagel lights upon the distinguishing characteristic of consciousness, no matter in what form: for an organism to have experience and to be conscious, there must be something it is like to be that organism, something it is like *for* that organism (392). Any analysis of mental phenomena would need to take into account this subjective character of experience. The problem being that for this reduction to be successful, “phenomenological features of the mind” must be given a physical account, which seems impossible, given that they are subjective – they cannot be separated from their single point of view. He illustrates his point with a thought experiment: can we (that is, human beings) know what it is like to be a bat?

The answer, predictably, is no. Nagel illustrates that it is impossible for us to know what it is like *for a bat*, to be a bat. It is structurally impossible for us to know how a bat experiences the world; our sensory apparatus is radically different from that of bats. Virtually experiencing the world-view of a bat or imagining yourself as a winged creature hanging upside down from the roof of a cave does not count. In such a case we will only know what it is like for us to be a bat, which is still not an objective account of “batness.”

With this Nagel comes to the crux of the matter: “the relation between facts on the one hand and conceptual schemes or systems of representation on the other” (396). Nagel declares his “realism about the subjective domain in all its forms” to lead him to the belief that there are facts that exist beyond the reach of human concepts (*ibid.*), for which we never will or could possess the concepts to represent or comprehend them. Nagel’s reflection on what it is like to be a bat leads him to conclude that: “there are facts that do not consist in the truth of propositions in the human language” (*ibid.*) For example, facts about what it is like for organisms other than oneself (especially organisms that are structurally very different from oneself) to be conscious.

The bearing of this on the mind-body problem is that the facts of experience (the facts of experiencing consciousness) are necessarily accessible from only one point of view. Given the subjective character of experience, Nagel sees it as a mystery how the *true* character of experience might be revealed in the physical operation of the organism (397). The reason he gives for his scepticism is that the physical operation of the organism is necessarily the domain of *objective facts* – objective facts being facts that can be observed and understood from many different points of view, including those of individuals with differing perceptual systems. Nagel’s definition of objective facts rests on the assumption that, the less our description depends on a particularly human point of view, the more objective it is (398). It is impossible to render an account of experience, without including the point of view of the “experiencer.” Because experience does not have an objective character, Nagel finds it difficult to understand how “a physiologist, or a Martian”, could study someone’s or something’s brain and observe their mental processes “from another point of view” (398). If experience is

essentially subjective, shifting to the objective viewpoint moves us away from the nature of the phenomenon. How can we, in this case, reduce our dependence on the individual, or species-specific point of view and direct our attention to the mental, as an object? How can the mental be “reduced” to the physical?

The argument is essentially that, seeing as conscious experience is necessarily subjective, it is not possible to explain it in objective terms. Ratifiable knowledge needs to be objective, and hence we cannot know what it is like for a bat, to be a bat. Mental states need to be known through the observation of physiological processes, and the species-specific viewpoint must be eliminated. Essentially, if mental processes are physical processes, Nagel argues that there must be something that it is like to be a physical process. He cannot see how this is possible, and as a result he concludes that, in order for us to have any legitimate knowledge of the mental, we need to develop an “objective phenomenology,” which does not rely on empathy or imagination. Through this method we should be able to describe subjective experiences to those who cannot experience them.

Hofstadter (1982:403-414) describes Nagel’s as an attempt to “subjectively know what it is objectively like” to be (409), and dismisses Nagel’s argument as an “over-facile thought experiment” (406). He argues that Nagel wants to use the verb “to be” in such a way that it will not refer to a particular subject, like a bat, but to that which is “subjectless,” referring to the “batness” that all bats have in common. As he puts it: “There is a be-ee here, without the be-er”(407).

Hofstadter believes that Nagel’s fundamental error in this essay is the assumption that to be justifiable, knowledge has to be objective. The contradiction in this argument is that, in order to describe consciousness in a legitimate way, we need to objectively describe what it is subjectively like to be (409). Hence, if experience or consciousness cannot be described in objective terms, any knowledge that we have thereof cannot be considered to be valid knowledge. Hofstadter disagrees with Nagel’s contention that we cannot know what it is like to be a bat, because of bats’ perceptual apparatus, or their “minds” are so vastly different from ours. On the contrary, he argues

“the modality of sensory input is quite interchangeable”.¹⁸⁷ The difference in “consciousness” between us and bats (or BATs)¹⁸⁸ has more to do with their much more limited range of conceptual and perceptual categories, along with the stress on things that are important in the life of a bat, than with the essential impenetrability of “batness” to the human mind.

Hofstadter goes on to highlight the role that *language* plays in our ability to exchange ideas and experiences. He argues (convincingly) that it is highly unlikely that bats wonder about, or exchange ideas about, what it is like to be a bat, or some other BAT. And the reason is, of course that bats do not, as far as we can tell, have language, at least not as far as a medium in terms of which ideas can be exchanged. While human beings do have the capacity to communicate on a sophisticated level through language (by which Hofstadter generally means some form of symbolic communication – he explicitly includes spoken language, movies and gestures). Hofstadter argues that humans have at their disposal various media that aid them in projecting, and thus absorbing and understanding (to a greater or a lesser degree) foreign points of view. To quote Hofstadter directly: “Through a universal currency, points of view have become more *modular*, more transferable, less personal and idiosyncratic” (413). The leap from our inability to penetrate the “batness” of being a bat, to concluding that we cannot know what it is like to be another person seems absurd. It stands to reason that the experiences of being human have enough common ground for such experiences to be transferable, to a significant degree at the very least, between different individuals.

Hofstadter does not see the singularity of the subjectiveness of consciousness, the inability to render an objective account of subjective experience, as an insurmountable obstacle in our ability to acquire knowledge about consciousness – even the consciousness of another subject. The reason for this is that Hofstadter does not make the same fundamental

¹⁸⁷ Consider the example (Hofstadter 1982:411) of the possibility of producing visual experiences in the blind as a result of tactile stimulation.

¹⁸⁸ Hofstadter names all sentient things, all things that it is something like to be, BATs (i.e. be-able things) (1982:409).

assumption that Nagel makes: he does not believe that knowledge is and should necessarily be objective. In fact, Hofstadter argues that knowledge is “a curious blend of subjective and objective” (413). It is in fact, the subjective aspect of knowledge and the fact that the subject can verbalise his/her subjective experiences in a medium of communication that enables us to share knowledge in the first place. We *can* experience what it is like to be or to do X, through a sequence of simulation processes, through language (414). The fact that such a simulation can never be the “original experience” is not an anomaly, but part of the structural precondition of the possibility to communicate knowledge at all!

Hofstadter draws, in part, on the work of Richard Dawkins, and notably on his concept of *memes*. Dawkin’s theory of memes ties in well with the aim of this chapter and we will briefly discuss his ideas before consolidating our themes and drawing this discussion to a close. We have seen that Dennett also draws upon Dawkins to develop his theory of heterophenomenology, and, as we will discuss, comes to roughly the same conclusions as those held forth by Hofstadter. By way of conclusion we will argue that the positions of both Dennett and Hofstadter that we have discussed in this work can serve to confirm and elaborate upon the implications of a complex model of the self as developed in the preceding chapter.

3. Culture red in tooth and claw?

Richard Dawkins devotes the last ten pages of his book *The Selfish Gene* (1976) to his new concept: *memes*. Memes are Dawkin’s cultural answer to genes. The gist of Dawkin’s biological argument is that all organisms (animals, plants, bacteria, viruses, etc.) are survival machines. All of these “machines” bear replicators¹⁸⁹: i.e. genes. Survival machines are carriers of copies of the same replicator – DNA molecules (23). Dawkins aims to explain the behaviour of organisms, particularly individual selfishness and individual altruism in terms of *gene selfishness*. A (admittedly oversimplified)

¹⁸⁹ Replicators are molecules that have the capacity to create copies of themselves (Dawkins 1976:16).

version of his argument runs as follows: in all likelihood molecules that could copy themselves – replicators – developed in the primordial soup by accident. Inevitably the primordial soup was incapable of supporting an infinite amount of replicators, and *competition* developed between the different replicators (20). Over time, some replicator varieties must have gone extinct, while surviving varieties had to struggle for existence. This struggle would have favoured any mutations that ensured greater stability in replicator varieties. Dawkins speculates that replicators had to construct containers for themselves, vehicles with which to ensure their continued existence. Consequently, he speculates, the first living cells developed. These primal living cells would have been the first examples of survival machines. As can be imagined, Dawkins envisions survival machines becoming bigger and more elaborate by means of a cumulative and progressive process (21). Four thousand million years later these ancient replicators:

...[S]warm in huge colonies, safe inside gigantic lumbering robots, sealed off from the outside world, communicating with it by tortuous indirect routes, manipulating it by remote control. They are in you and in me, they created us, body and mind; and their preservation is the ultimate rationale for our existence. They have come a long way, those replicators. Now they go by the name of genes, and we are their survival machines (1976:21).

An important point that Dawkins tries to bring across with this image is that genes indirectly control the manufacture of bodies. In his own words: “The body is the genes’ way of preserving the body unaltered” (barring mutations, of course)(25). Although essential to survival, an organism’s acquired characteristics are not inherited by its offspring. In other words, knowledge and wisdom acquired by an organism through a lifetime cannot, by any genetic means, be passed on to its offspring; each new generation starts from scratch (25). Or does it?

Dawkins is careful to emphasise that genes do not have foresight. They are not conscious entities, organising into structures in order to fulfil some kind of teleological master plan. The process of replication happens

blindly; a process of automatic selection between molecules, based on their fecundity, longevity and copying-fidelity (25).

Dawkins notes that man is unusual with regard to this scenario in one aspect: that of culture. He develops a theory of culture in which culture itself can undergo a form of evolution, through a process of cultural transmission instead of genetic transmission.¹⁹⁰ Furthermore, cultural evolution is orders of magnitude faster than genetic evolution (203). Dawkins names instances of cultural products that are transmitted and inevitably undergo transformation such as language, dress, diet, ceremonies, art, customs, architecture, and technology. All of these aspects of human life change over historical time, in such a manner as to resemble a genetic evolution that has been speeded up (204). Surprisingly though, Dawkins argues that, in order to understand the evolution of modern man, one needs to discard the gene as the sole basis of our ideas on evolution. In fact, he likens the current state of human culture to the primeval soup and names a basic replicatory cultural entity: the meme. He derives his new noun from the Greek root *mimesis*. A meme is a unit of cultural transmission, a unit of *imitation*. Among the examples of memes that Dawkins lists are the following: tunes, ideas, catch-phrases, clothes fashions, ways of fashioning pots or building arches (206). Memes do not propagate themselves by perpetually leaping from parental body to offspring, but by leaping from brain to brain. Dawkins considers memes to be living structures, not just metaphorically, but technically as well (207). This image of memes allows him to envision the brain as a vehicle that is “parasitised” by memes, similar to the way in which a virus may parasitise a host cell.

The way that memes replicate themselves and achieve transmission between “host brains” is through a process of *imitation*. An idea, like that of an afterlife, for instance, is replicated “through the spoken and the written word, aided by great music and art”(207). Dawkin’s example of great music and art is, of course, just a fraction of the imitative processes that can be used to spread ideas. For a start, bad music and art are replicators that can be just as

¹⁹⁰ Dawkins does concede that cultural transmission is not unique to man (Dawkins 1976:203). Presumably cultural transmission is at its most pronounced in the activities of man.

efficient in spreading ideas (have “survival value” in Dawkins’s terms), if not more so in some cases.

Dawkins attributes the survival value of a meme to its psychological appeal in a given cultural environment. Memes that have psychological appeal are copied through successive generations’ brains, regardless of their accuracy or effectiveness (Dawkins uses the example of the “god-meme” to substantiate this position). In other words, just as with genetic evolution, some memes are more effective than others when it comes to survival. Dawkins believes that successful memes have the same attributes as successful genes: longevity, fecundity, and copying-fidelity. In terms of fecundity and longevity, the analogy between memes and genes seems relatively unproblematic. Some memes spread like wildfire, but their popularity is short-lived (think of the majority of songs on the pop-charts or the latest fitness fad), while others have been around for thousands of years and are likely to be around for some time to come (the various major world religions, for instance).

Dawkins does admit, however, that his analogy seems to break down somewhat when it comes to copying-fidelity. Memes do not seem to be high-fidelity replicators at all, especially when compared with the relative stability found in genes when it comes to the copying process. Memes appear to be a lot more mutable than genes, endowing them with a certain fluidity, with more “licence”, if you will, when it comes to imitation. In other words, that meme transmission is subject to a kind of continuous mutation. Surprisingly perhaps, in an attempt to address this apparent discrepancy, Dawkins revises some of the assumptions that have been made about genes. He concludes that the possibility does exist that the analogy does not break down, if genes are less particulate than they have portrayed as thus far. In this respect, in the case of genetically inherited characteristics there are so many genes involved that genes *seem* to “blend.” Dawkins does *not* mean that genes are not particulate, but he does concede that it becomes very difficult to define what constitutes a gene. Dawkins settles the question in the following manner:

The “gene” [is] defined, not in a rigid all-or-one way, but as a unit of convenience, a length of chromosome with just enough

sufficient copying-fidelity to serve as a viable unit of natural selection (210).

In a similar manner it is difficult to determine what a single, viable, meme-unit might be. Dawkins settles for a single unit, or phrase, or word etc. that is sufficiently distinctive and memorable to be abstracted from its context and – while still being recognised as that particular, distinctive unit – be used in another context (210).

Just as genes are not purposeful agents, memes do not operate with any foresight or agenda (211). Meme selection is a blind process. As to the competitive aspect of gene selection, Dawkins finds a way to stretch his analogy to incorporate a form of “competitiveness” into mimetic evolution. Instead of biological resources, time and storage space are the most important limiting factors in the human brain. In order to survive a meme needs to attract attention to itself, at the cost of its “rival” memes. Memes that have a greater psychological impact have a better chance to receive attention and be perpetuated. Dawkins also mentions other commodities for which memes compete: television time, newspaper space, billboard space, etc. All of these are means of transmission, ways of influencing other human beings and effecting changes in their brains. And, just as genes evolve into evolutionary stable sets of genes, memes evolve into evolutionary stable sets of memes that reinforce one another – a particular culture, for example.

In short, Dawkins concludes that:

Once genes have provided their survival machines with brains which are capable of rapid imitation, the memes will automatically take over. We do not even have to posit a genetic advantage in imitation, though that would certainly help. All that is necessary is that the brain should be capable of imitation: memes will then evolve which exploit the capacity to the full ... We are built as gene-machines and cultured as meme-machines...(215).

A significant aspect of the culturing process is, of course, the structure of the human brain, and consciousness in particular. Dennett is worth quoting at length in this context:

Since this new machine [consciousness] created in us is a highly replicated meme-complex, we may ask to what it owes its replicative success. We should bear in mind, of course, that that it might not be good for anything – except replicating. It might be a software virus, which readily parasitises human brains without actually giving the human beings whose brains it infests any advantage over the competition. More plausibly, certain features of the machine might be parasites, which exist only because they can, and because it is not possible – or worth the trouble – to get rid of them. William James thought it would be absurd to suppose that the most astonishing thing we know of in the universe – consciousness – is a mere artefact, playing no essential role in how our brains work, but however unlikely it might be, it is not entirely out of the question, and hence not really absurd. There is plenty of evidence around about the benefits consciousness apparently provides us, so we can no doubt satisfy ourselves about its various *reasons d'être*, but we are apt to misread that evidence if we think that a mystery remains unless every single feature has – or once had – a function (from our point of view as consciousness-“users”) (Dennett 1991:221).

Dennett is careful to point out that evolution is not teleological. Consciousness did not develop “for” anything. Most probably it endows the organism that possesses it with certain evolutionary advantages. If not, it will quite probably go the way of other failed evolutionary experiments. Important to the current discussion is the idea of human consciousness as a complex of memes, or as Dennett puts it, meme-effects on the brain (1991:210). Also of great importance is Dawkins’s description of culture as a medium for transferring memes between brains.

Given that consciousness only arises as a result of interaction with other human beings – as a result of culture – the question arises as to what culture is. In the light of the preceding discussion “culture” seems to be the collective of manifestations of existing memes. Any means of replicating memes (communicating) hence becomes a cultural enterprise. In the previous chapter we saw that culture takes the part of the environment when it comes to seeing the self as developing in a complex system. We found that consciousness and a sense of self are impossible without input from the environment. In this respect, the view that Dawkins, Dennett and Hofstadter take on culture accords very well our view of culture as part of a complex system in which the self emerges. In this respect it seems likely that we have to re-evaluate the role that culture has traditionally been accorded in philosophical theory. The peripheral role that has traditionally been granted to cultural activities in the past can explain Nagel's profound scepticism when it comes to the possibility of people trading ratifiable knowledge about being. The suspicion with which he views subjective accounts and the interpretation of those accounts points to a profoundly sceptical approach to language and culture that has been strongly rooted in the Western philosophical tradition from the first.

3. Plato's view on “art”

The conception of cultural phenomena as peripheral to *essential* activities – such as gathering food, perhaps? – that presumably do more to contribute to the survival of the species, stems from conceptions of certain aspects of culture, especially artistic activity, that are, in part, a legacy of Plato's views on art.¹⁹¹ “Art” here should be understood as a shorthand term for most cultural activities – all manifestations of *meme-transference*. When it comes to discussing human culture it becomes exceptionally difficult to accord different practises to different categories. For the purposes of this short

¹⁹¹ Gardener (1996:250), who cites art's “essential connections with pleasure, play and imagination and its freedom from reason and practical purposes” as positive reasons for scepticism about the value of art, sums up such a conception of art.

chapter we will ignore these complexities, so as to make a general point about culture, in a very broad sense of the word.

Crudely Plato's conception of art runs something as follows: artists copy the world in various media, generally for their own amusement and the entertainment of those who busy themselves with such matters. Artistic activities are at best frivolous and at worst geared towards trying to undermine the accepted moral and political practices of a given society (Plato 1981: 421-439). Are cultural products, meme-complexes, activities that are not only peripheral amusements but that also do not provide any practical contribution to humanity's general welfare?¹⁹²

To the ancient Greeks, especially Plato and Aristotle, art is not "real" in the same way that the world is real (Eaton 1988:92).¹⁹³ With his "conception of the hierarchical structure of reality" (Verdenius 1971:268), Plato differentiates between the visible realm and the intelligible realm. The visible realm is the world, as we know it – the empirical reality that all people have access to. The intelligible realm is the absolute form of Good and "responsible for everything right and good" – it is the source of all light, being, reality and intelligence. Empirical reality is a descent from the Divine realm, and consequently an imperfect copy of the ideal.

There are different planes of being which each aspires to express the values of the one superior to it, except for the Good, which is "absolutely real" (Verdenius 1971:268). Verdenius goes on to explain that the *degree* of reality of a realm is dependent on its degree of approximation to the Ideal realm, what he calls "eternal Being". The empirical world is not true reality, but an

¹⁹² Nelson Goodman also asserts that the (mis)conception of art as frivolous, as entertainment, has done a disservice to contemporary views on the arts and on art education:

"Serious study of education for the arts has also been stunted and sidetracked by the prevalent notion that the arts are merely instruments of entertainment. Some newspapers list plays, concerts, and exhibitions under "amusements"; and among a week's amusements may be a Bach Mass, *King Lear*, and an exhibition of Goya's *Disasters of War*. No real progress in attitudes toward education can be hoped for when Cézanne's pictures are classed with cookouts, and arts programmes with playgrounds. On the other hand, we encounter almost as often the equally detrimental mistake of exalting the arts to a plane far above most human activities, accessible only to an elite" (1984: 154).

¹⁹³ Where Plato sees this as reason enough to reject art, Aristotle ascribes a more positive role to artworks: they are a source of knowledge (Eaton 1988:22).

approximation to it, a copy that strives to the Good, but falls short. Art is yet another plane removed from the Good. According to Plato an artistic rendering of the empirical world is inevitably an imperfect copy. An artist imperfectly copies an empirical world, which is in itself an imperfect copy of “true” reality (Plato’s Ideal realm), and thus renders a phenomenon which, according to Plato’s scheme, is “a third removed from reality” (Plato1981: 425). As far as degrees of reality go, Plato seems to regard works of art as scraping the bottom of the barrel. Pictures and poems and such like are secondary phenomena and tell us *nothing* about life. Consequently Plato believes that art has no value, and not only that, it can be an undermining and dangerous activity to boot.

As, so Plato’s argument goes, an artist does not need knowledge of reality to render a superficial representation of it, (s)he is capable of rendering something without the correct opinion about the “goodness” or “badness” of the thing (s)he is representing (430). Thus, not only does art not have any serious value but, seeing that it is not a product of reason, it has a low degree of truth *and* it appeals to “the lower elements in the mind,” possessing the threat of ruining the higher elements. Art carries the danger of obscuring or even distorting the truth, and thereby having a negative influence on those exposed to it. Plato gives poetry “the terrible power to corrupt even the best characters”, and concludes that art should not be allowed in his ideal state.

With the emergence of popular culture, mass communication and a media literate global population, another realm is added to Plato’s hierarchy. Or, perhaps, it can be integrated into the realm of “art” to the extent that increased technical ability to produce and distribute communicative signs has exaggerated the “logic” that characterises Plato’s artistic activity. Whereas Plato believed that art (understood as manifestations of meme-complexes) does not add to our knowledge of the world, the same cannot be said about the media – without access to the divergent forms of the mass media, we would have a vastly different picture of the world.

In the same vein, Plato’s contention that art is at best a meaningless activity, and at worst has the potential to corrupt characters contains a degree of contradiction. How can an activity that has no value and adds nothing to the

world have the ability to affect characters? It seems that, while Plato contends that art has no value, he does not believe it to be without any consequence. Artistic activity can have some very real effects. In the light of this Plato makes a concession to the “telling of tales” that makes use of indirect speech, in so far as tales have some pedagogical value (Ijsseling 1997:12).

In more recent times increased technical capability has provided a highly efficient means of reaching large numbers of people. The Nazi, Fascist and a succession of totalitarian regimes, as well as advertising companies, have all learned to harness a uniquely modern manifestation of Plato’s “telling of tales”: mass communication. In Germany in the 1930’s there was a conscious and deliberate attempt by the Nazi party to establish official Nazi ideology in all areas of culture and art and to eradicate alternative ideologies (Strinati 1995:5). “The aim was to enlist the help of intellectuals, writers, poets, painters, sculptors, musicians, academics, architects, etc., in order to establish Nazi ideology as Nazi aesthetics” (7). Both the Nazi’s and Plato realised that “art” was by no means a neutral activity and that it has the very real potential to shape public opinion and undermine prevailing political ideology.

From Plato’s criticism of art stems many of the subsequent views on, and criticisms of, art/culture in the Western philosophical tradition. Some contemporary theories on cultural products still rely on the basic presupposition that leads Plato to ban artists from his ideal republic: that artworks are representational or *mimetic*, and an imperfect representation at that.¹⁹⁴ Presumably there are better ways to render and communicate knowledge. In copying reality, art/culture runs the risk of getting it wrong, of not capturing and communicating the essence of the original on which it is based. Nagel’s criticism that we cannot rely on subjective accounts of experience, because they might not capture the essence of a particular “be-ee” (to use Hofstadter’s term) and might be misinterpreted, seems to be a variation on Plato’s scepticism. Art/ language/ human artefacts/ meme-

¹⁹⁴ Although a Neoplatonist, in turn, would attribute value to art in accordance with its ability to “purify” the world of matter and render images as close to the ideal as possible (Gombrich 1968: 133).

manifestations run the risk of being corrupted by *subjectivity* and must therefore be subjected to strict guidelines according to which they can be ratified – scientific principles. Unfortunately some phenomena do not lend themselves to be scientifically studied (consciousness, for example) and should therefore be banned, or at least considered to be impenetrable.

We will briefly look at an attempt by Popper to disperse some of the assumptions implicit in this position, with a view to render these “impenetrable” subjects open to objective study. As we shall see, his attempt is not at all convincing and we will argue that such a delineated approach to the intricacies of mind, culture and the physical world is too simplified and rigid to capture the interrelatedness that exists in the system world + consciousness + memes. In the final analysis Popper fails to account for the fundamental characteristic of consciousness and knowledge – what Hofstadter calls its “curious blend of subjective and objective” (1982:413).

5. Popper's World 3

Popper's (1977) approach is that of a philosopher of science, trying to account for the reality of “mental states” and their ability to affect the physical world. To this end, Popper busies himself with the status of ideas, wishes and thoughts as manifested in the empirical world through our representation of them in diverse signs and symbols. Theorising against an intellectual background where materialism is predominant and the physical sciences are the standard measure for validating theories, Popper tries to account for the “reality” of mental states and ideas, and in so doing to refute *radical* materialist theories that relegate ideas and mental states to insignificant epiphenomena of chemical and physical processes in the brain. He also opposes the idea that there is something unique to consciousness that makes it a somehow impenetrable and mysterious phenomenon. Popper supports his contention that mental states are real in the same way that physical states are real through recourse to *interaction*, where mental states are as real as physical states, because of their ability to interact with the physical world and cause tangible effects in that world. He also introduces interactionism as

answer to the mind-body problem through a tripartite division of the world into: a universe of physical entities – World 1; the world of mental states – World 2, and the products of the human mind – World 3.¹⁹⁵ As examples of World 3 objects he cites stories, explanatory myths, tools, scientific theories (whether true or false), scientific problems, social institutions and works of art¹⁹⁶ (38). These are all objects of our own making, but not necessarily the creations of individual people. Objects from World 3 are manifested in World 1 in the form of material objects (sculptures, paintings, books – whether scientific or literary).

But World 3 objects are not only “real” in their World 1 manifestations, but also in World 3 aspects by virtue of their interaction with World 1. For example, a particular sculpture can influence other sculptors on a World 2 level and then find new manifestations in subsequent World 1 products. In an attempt to pre-empt criticism that he is just reformulating a case of imitation, Popper uses the example of the production of a scientific theory: the scientist starts from a problem – a demanding intellectual task – which finds its origin in other theories encountered in the subject literature of his/her field (World 3 objects – theories – manifested in World 1 objects). The scientist then proceeds with a *creative effort*: that of trying to grasp and formulate an abstract problem. If successful, (s)he produces a solution (a new theory) which is then put into linguistic form and subjected to literary discussion and modification, and accepted or rejected. An important aspect of theories, according to Popper, is that despite their being products of human thought they have a measure of autonomy – they may have consequences that had

¹⁹⁵ Although this division can be criticised in many aspects, especially from the point of view of complexity theory, it is useful for the purposes of the current argument, in that it not only highlights the effects of human ideas on the physical world, it also places the various human disciplines of thought on an equal footing, with regard to their ability to interact with and influence the empirical world.

¹⁹⁶ Popper believes world 3 to be contingent, but insists that his is not a plea for relativism: “Human thought in general, and science in particular, are products of human history. They are, therefore, dependent on many accidents: “had our history been different, our present thinking and our present science (if any) would be different” (148). But, Popper continues, we can learn from our mistakes and science can progress, our theories are not arbitrary: “What is important is that there is no self-contradiction whatever in describing *scientific knowledge* or, say, *historical knowledge*, as consisting largely or wholly of hypotheses or conjectures, rather than as a body of known and well-established truths”(123).

not been foreseen when they were being created, and in this sense aspects of these theories can be said to have been discovered and objective. Thus World 3 can be said to be man-made in its origin and mimetic to the extent that new theories, etc. rely on theories, insights and accepted formulae preceding them. But, once theories exist, they take on a life of their own, producing new consequences and new problems.

From this, and the creative aspect of producing new scientific theories, Popper deduces the *objectivity* of World 3. Furthermore he deduces the World 3 status of science, and the influence of scientific theories on the world, which establishes the reality of objects in World 3 – including ideas, thoughts and wishes.¹⁹⁷ Important for our discussion is that Popper places artworks and scientific theories on a par as objects that make up World 3 – objects that both shape and are shaped by human thought, and which accord us access to Worlds 1 and 2. Both disciplines rely on a combination of the constraints and insights provided by their tradition and on creativity.

Language is a World 3 object, and can accordingly be endowed with objective status, as well as being considered to be “real”.¹⁹⁸ To Popper, the social character of language, and the fact that we can speak about other people and understand them when they speak about themselves, ensures that we are not only subjects, or centres of action, but can also be objects of our own critical thought and judgement – one cannot be fully human without mastering language (144) – we owe our humanity, our rationality to language, and thus to other people. We are products of World 3, which is in itself a product of countless human minds. Popper explicitly equates the process of acquiring World 3 objects with the process of acquiring a self: “One learns not

¹⁹⁷ Popper consigns artworks to World 3. Their world 1 manifestations have their origin in World 3 ideas and theories. World 3 objects do not lie in some ideal realm, independent of human thought and waiting to be discovered, but are the product of, and produce in their own right, a dynamic interaction between “ideas” and “reality”.

¹⁹⁸ Popper argues that the most fundamental of the World 3 learning processes is that of acquiring language. Language is what enables us to see and interact with the world. The physicist might primarily be interested in World 1, but in order to learn about World 1 (s)he needs to theorise, and for that (s)he needs to use objects from World 3 – primarily language – as tools. The physical scientist who studies World 1 has a vested interest in World 3 tools and their logical consequences, which are a prerequisite for doing “applied science” and using World 3 insights to affect World 1.

only to perceive and to interpret one's perceptions, but also to be a person and to be a self"(49). Without social interaction, Popper contends, we cannot develop a sense of self (111). And, of course, language in its various forms enables social interaction.

It might be tempting to conclude that, if we establish the "reality" of phenomena in terms of their ability to influence the empirical world (World 1), then, in terms of Popper's scheme, Plato already gives us reason enough to postulate art/culture as a "real" entity by admitting that it has the potential to corrupt characters and subvert prevailing moral standards. If art is the concrete manifestation of ideas, thoughts and wishes and this concrete manifestation has the ability to affect the world artistic activities become primary rather than secondary activities, in that they serve establish ideas about the world that would not have been accessible in any other way. It is doubtful, however, that Popper's attempt to establish the objective nature of language is what Nagel has in mind when he argues that we need to formulate an objective phenomenology in terms of which we can theorise on consciousness. Presumably Nagel would argue that even language as an "objective medium" does not guarantee that one can gain insight into what it is like to be another person, or a bat for that matter, because consciousness is necessarily subjective. Is it possible that Hofstadter is right in asserting that knowledge is a blend of objectivity and subjectivity, and that it is the subjective aspect of knowledge and the fact that the subject can verbalise his/her subjective experiences in a medium of communication, that enables us to share knowledge at all?

At this juncture it seems that Dennett's theories holds the potential solution to the problem of the "mysterious" aspect of consciousness. Dennett's proposed heterophenomenology seems to be the most viable suggestion of those that we have discussed on how to conduct an objective study of what seems to be an inherently subjective phenomenon. Such a heterophenomenology could be a successful tool in terms of which to explore consciousness, because it applies to one's own consciousness as well. In chapter 3 we have seen that Dennett argues that even one's own consciousness of self is a narrative fiction: the result of streams of multiple

possible narrative drafts of self and consciousness that are to be found in the unconscious. This fiction is an account of all that the subject sincerely believes to exist in his/her conscious experience. Dennett insists that such a narrative *is* a portrayal of exactly what it is like to be that subject – that subject's experience – and is an adequate basis from which to explore this heterophenomenology. Dennett's approach has much in common with that of Popper. Consider, for example the following quotation, already discussed in chapter 4, but which merits repetition here:

The heterophenomenology exists – just as uncontroversially as novels and other fictions exist. People do undoubtedly believe they have mental images, pains, perceptual experiences and all the rest, and *these* facts – the facts about what people believe, and report when they express their beliefs – are phenomena any scientific theory of the mind must account for. We organise our data regarding these phenomena into theorist's fictions, “intentional objects” in heterophenomenological worlds. Then the question of whether items thus portrayed exist as real objects, events, and states in the brain – or in the soul, for that matter – is an empirical matter to investigate (1991:98).

Dennett does not, however, try to account for the “reality” of mental states. Nor does he create a phenomenological gap between mental states and the physical world, endowing the former with the ability to affect the latter. Dennett recognises the reality of “mental states” as a matter of some complexity. He emphasises that a conscious mind is an observer, and that where there is a mind, there is a *point of view*. He also provides us (with the help of Dawkins) with a much more dynamic picture of how consciousness/self is formed and maintained, *and* of the potential that it has to evolve over time within a single individual, and from generation to generation. Dennett insists on the material character of consciousness and the self, but refrains from creating the impression that distinct mental units exist that are somehow represented in the brain. Dennett's description of the “distributed content-discriminations” within the brain, with the possibility to

become something like a narrative stream, subject to continual “editing” by many processes distributed in the brain, is an ideal accompaniment to a complex theory of self. Especially given the emphasis that is placed on the role of the environment in structuring the self, and the need it creates to describe the structure of such an environment. With the following statement Dennett provides us with the last of the tools necessary to compile a model of the uniquely human environment in which the self emerges:

Human consciousness is *itself* a huge complex of memes (or more exactly, meme-effects in the brain) that can best be understood as the operation of a “*von Neumannesque*” virtual machine¹⁹⁹, *implemented* in the *parallel architecture* of a brain that was not designed for any such activities (210).²⁰⁰

The idea that a narrative fiction is as real as story-telling, for example, novels and other forms of language manifestation (or meme-manifestations, if you will), coupled with the idea of consciousness being a complex of meme-effects in the brain creates a unique role for culture, as environment, in the development and structure of the self, as well as the unique character of the human environment.

6. The complex self and the changing role of a uniquely human environment

In the previous chapter we argued that culture provides us with permanent memory-traces. We also emphasised the importance of other “mature human brains” to use Trevarthen’s phrase, and the cultural entities through which different brains communicate with one another. In his discussion on Dawkin’s memes Dennett has already pointed out the vital role that the culture in which they manifest themselves plays in the evolution of

¹⁹⁹ I.e. a computer with a fixed (hardware) structure that can run different kinds of soft-ware, and as such can function as a series of different “machines” with divergent capabilities (see Chapter 2).

²⁰⁰ See footnote 155.

consciousness. And we proposed to argue that given our conception of the self as an emergent property in a complex system, of which culture (as environment) makes up a great part, that we should rethink many of our presuppositions about the cultural phenomena that we create and endorse.

The complex self is an emergent property of the complex system composed by human genetic predisposition to acquire or rather manufacture a self, and the environment from which the elements with which a self is constructed can be gathered. Another important factor in the development of a complex self is the history of that self, which is the recorded memory-traces of past experiences. We have seen that the human environment is unique, in that it is the product of many consciousnesses, embedded in concrete cultural phenomena. Dawkins gave us the term memes with which we could roughly delineate “units” of cultural phenomena. In this final chapter we have seen that consciousness is not only something that we *can* attain knowledge about, the structural characteristics of consciousness that make this attainment possible are also the characteristics that are responsible for forming consciousness. Not only can we talk about consciousness, the manifestations of consciousness in culture, memes, are involved in structuring both consciousness and self.

A self, as an emergent property of the complex system human mental apparatus + environment/culture is the conscious sense of self, the conscious model of itself that the human organism needs in order for it to function successfully. This model that a human being might have of him/herself can be radically contingent, in accordance with the cultural factors that a person is exposed to. The model is also constrained by factors that we have likened to attractors within a complex system. Given certain possible variations from person to person, the mental apparatus, at base, seems to be genetically wired in a relatively stable pattern. Genetic changes will occur over time, of course, but we would argue that most *homo sapiens*, even over time, have enough characteristics in common, that given the ability to simulate experience through language, they can understand what it is like for a particular person, to be that person.

By the same token, if the medium of simulation, language in its various forms, does play such a significant part in the development of the human mind as Freud, Dennett, Trevarthen and even Popper contend, it stands to reason that what “developing young brains” are exposed to, should be brought under careful consideration. If one’s brain is structured in part by the meme-effects of only those memes that one is exposed to, it stands to reason that what would emerge as one’s self would be significantly constrained by the memes that one has been exposed to during one’s lifetime. If these memes are all of a particular sort, inaccurate or not conducive to the formation of a self that is an asset to one’s ability to live well, the resultant sense of self might become a liability. So much so that it might inhibit a person’s ability to project possibilities accurately and factor its own nature into decisions to such an extent that it manages to live successfully. Given the non-linearity of a complex system, nothing that we are exposed to can reliably be expected to be of no consequence. In the light of this it seems that we should be very careful as to what memes we are exposed to, and what we expose others to. It seems that the self is contingent in more ways than one. Therein lies both its greatest strength: the ability to adapt and learn, and its greatest weakness: the possibility that it can be completely erroneous in conception and hence ineffectual or even counterproductive. The complex environment, consisting of various cultural entities and of people, seems to be the single greatest determining factor in how our self will be, and as such can be either munificent or treacherous. Most probably it is an uneasy mixture of the two.

Postscript

By way of conclusion we will give a short summary of the issues raised, the arguments given, and the conclusions drawn in the preceding pages. At issue here was the status of the self in its contemporary context. The question arose as to the viability of some of the current assumptions about the self, as well as possible origins of these assumptions. It was found that many of the current preconceptions of the self had their origin in Cartesian dualism and the subsequent centuries-long debate that the work of Descartes had inspired. A central conception of the self to emerge from this debate is that of the self as an immaterial entity, whether wholly independent from, or partly connected to, the material body. This conception of the self, among other things, postulates the self as *not* subject to the processes of causality operating in the material world. In this manner the concept of the free will of the subject can be accommodated into a causal world-view, ostensibly without upsetting the principles of either concepts.

We proceeded to criticise the idea of the immaterial self as unfounded and insufficient to the task of accounting for the formation and structure of the self. Freud, in postulating the unconscious, opened up the possibility of accounting for the material nature of a contingent self, bound both by its body and its environment. Far from being an essence, the rational capacity of man, for instance, the self becomes a complex function of brain processes. With the advent of both psychology and neurology, rooting our theories of the self in the empirical and the material becomes not only a viable option, but seemingly inevitable. Summarily conceiving of the self as essentially non-material when the possibility exists that the phenomenon can be subject to empirical scrutiny and scientific research, in keeping with the principles of both neurology and psychology, becomes indefensible. With this in mind we discussed the work of Daniel Dennett, a philosopher very much concerned with the self as material phenomenon. Accordingly, Dennett emphasises the role that evolution plays in the formation of the self. We saw that, not only can the existence of the self be accounted for as a product of an evolutionary process, where an organism that can factor a realistic conception of itself into its survival strategies would have a distinct advantage over organisms unable

to do this, but that the human self is part and parcel of a relatively new evolutionary medium, presumably unique to man, that of the cerebral cortex and its resultant consciousness. Suddenly the self becomes a construct, geared towards the survival of the species, and limited by the possibilities of a specific mental apparatus in a specific environment.

Furthermore the possibility was raised that insights gained from a distinctively contemporary discipline, that of complexity theory, could be used to construct a viable and useful model of this “new”, material self that does not need to fall back on unfounded conceptions of the immateriality of the self. It was argued that the self, as an emergent property of the interaction between the mental apparatus and the environment, would be subject to many complexities and intricacies, characteristics that would be best served by a model that is based upon the principles of the complexity sciences. The application possibilities of such a model was discussed in some detail in chapter 4 and it was concluded that not only would such a complex model of self be viable, it might also be useful in highlighting some important aspects that feature in the formation and maintenance of a effective and healthy self. Of particular importance in this regard is the role that the environment plays in the formation of the self.

One interesting question that has been raised by this discussion, but which has not received sufficient attention here is the implication that a theory of a complex, material self would have on questions of free will and determinism. Unfortunately an exploration of these issues will stretch far beyond the scope of this work. We will have to content ourselves with mentioning them as interesting and challenging issues that result out of our discussion that will hopefully become the subject of study in the future.

Bibliography

'gren, H. 1998: "Chinese Ideas of Mind"

In *The Oxford Companion to the Mind*. R. L. Gregory (ed.). Oxford: Oxford University Press.

Armstrong, A. H. 1972: *An Introduction to Ancient Philosophy*. London: Methuen and Co Ltd.

Armstrong, D. M. 1998: "Mind-body Problem, Philosophical Theories"

In *The Oxford Companion to the Mind*. R. L. Gregory (ed.). Oxford: Oxford University Press.

Auyang, S. Y. 1998: *Foundations of Complex-System Theories in Economics, Evolutionary Biology, and Statistical Physics*. Cambridge: Cambridge University Press.

Ayer, A. J. (1998): "Mind and Body"

In *The Oxford Companion to the Mind*. R. L. Gregory (ed.). Oxford: Oxford University Press.

Bak, P. 1997. *How Nature Works. The Science of Self-Organised Criticality*. Oxford: Oxford University Press.

Bateson, G. 1972. *Steps to an Ecology of Mind*. New York: Ballantine Books

Cilliers, P. 1989. "Brain, Language and Consciousness. A Post-Structural Neuropsychology". Thesis delivered for the degree of Master of the

Arts at the University of Stellenbosch.

Cilliers, P. 1998. *Complexity and Postmodernism: Understanding Complex Systems*. Routledge, London and New York.

Cilliers, P. 2000a. "What Can We Learn From a Theory of Complexity?"
in: *Emergence*. 2(1), 23-33

Cilliers, P. 2000b. "Rules and Complex Systems."
in: *Emergence*. 2(3), 40-50

Cilliers, P. and T. de Villiers (2000) "The Complex 'I'"
in Wendy Wheeler (ed.) *The Political Self*. London: Lawrence and Wishart.

Cilliers, P. 2001: "Boundaries, Hierarchies and Networks in Complex Systems"
in: *International Journal of Innovation and Management*, Vol. 5, No. 2
(June 2001) pp. 135-147

Cilliers, P. and Gouws, A. 2001: "Freud's 'Project', Distributed Systems, and Solipsism" *South African Journal of Philosophy*, 2001, 20(3).

Cornell, D: *The Philosophy of the Limit*. New York and London: Routledge.

Culler, J. 1994: *On Deconstruction. Theory and Criticism after Structuralism*. London: Routledge.

Dawkins, R. 1976: *The Selfish Gene*. Oxford: Oxford University Press.

Degenaar, J. J. 1993: "Art and culture in a changing South Africa" *South African Journal of Philosophy*, 1993, 12(3).

Dennett, D. C. 1991. *Consciousness Explained*. London: Penguin.

Dennett, D.C. 1998. "Consciousness"

In *The Oxford Companion to the Mind*. R.L. Gregory (ed.). Oxford: Oxford University Press.

Derrida, J. 1978: "Freud and the Scene of Writing"

In: *Writing and Difference*. Chicago: Chicago University Press.

Descartes, R. 1978: *A Discourse on Method, Meditations, and Principles*. London: Everyman's Library.

Eaton, M. M. (1988): *Basic Issues in Aesthetics*. California: Wadsworth Publishing Company.

Fischer, E. 1959: *The Necessity of Art. A Marxist Approach*. Harmondsworth: Penguin Books Ltd.

Flanagan, O. 1991: *The Science of Mind*. Cambridge, Mass.: MIT Press.

Freud, S. 1901: *The Interpretation of Dreams*. Standard Edition, Volumes 4-5, London: Hogarth Press and the Institute of Psycho-Analysis.

Freud, S. 1911: *Formulations on the Two Principles of Mental Functioning*. Standard Edition, volume 12, London: Hogarth Press and the Institute of Psycho-Analysis.

Freud, S. 1915a: *Instincts and Their Vicissitudes*. Standard Edition, volume 14, London: Hogarth Press and the Institute of Psycho-Analysis.

Freud, S. 1915b: *Repression*. Standard Edition, volume 14, London: Hogarth Press and the Institute of Psycho-Analysis.

Freud, S. 1915c: *The Unconscious*. Standard Edition, volume 14, London: Hogarth Press and the Institute of Psycho-Analysis.

Freud, S. 1917: *A Metapsychological Supplement to the Theory of Dreams*. Standard Edition, volume 14, London: Hogarth Press and the Institute of Psycho-Analysis.

Freud, S. 1920: *Beyond the Pleasure Principle*. Standard Edition, volume 18, London: Hogarth Press and the Institute of Psycho-Analysis.

Freud, S. 1923: *The Ego and the Id*. Standard Edition, volume 19, London: Hogarth Press and the Institute of Psycho-Analysis.

Freud, S. 1925: *A Note upon the "Mystic Writing-Pad."* Standard Edition, volume 19, London: Hogarth Press and the Institute of Psycho-Analysis.

Freud, S. 1950 [1895]: *Project for a Scientific Psychology*. Standard Edition, volume 1, London: Hogarth Press and the Institute of Psycho-Analysis.

Gardener, S. 1996: "Aesthetics" in

Bunnin, N. and E.P. Tsui-James (eds): *The Blackwell Companion to Philosophy*. Oxford: Blackwell.

Gaze, R.M. and Taylor, J.S.H. 1998: "Neuronal Connectivity and Brain Function"

in *The Oxford Companion to the Mind*. R. L. Gregory (ed.). Oxford: Oxford University Press.

Gell-Mann, M. 1994: *The Quark and the Jaguar. Adventures in the Simple and the Complex*. London: Little Brown Company.

Gombrich, E. H. (1968): *Art and Illusion. A Study in the Psychology of Pictorial Representation*. London: Phaidon Press.

Goodman, N. 1984: *Of Mind and Other Matters*. Cambridge, Mass.: Harvard University Press.

Guyer, P. (ed) 1995: *The Cambridge Companion to Kant*. Cambridge: Cambridge University Press

Gregory, R. L. 1998a "'Emergence' and 'Reduction' in Explanations"

in *The Oxford Companion to the Mind*. R. L. Gregory (ed.). Oxford: Oxford University Press.

Gregory, R. L. 1998b "Turing",

in *The Oxford Companion to the Mind*. R. L. Gregory (ed.). Oxford:
Oxford University Press.

Hatfield, G 1995 "Empirical, Rational, and Transcendental Psychology:
Psychology as Science and Philosophy"

in: Guyer, P. (ed) 1995: *The Cambridge Companion to Kant*.
Cambridge: Cambridge University Press

Hofstadter, D.R. and D. C. Dennet 1982: *The Mind's I. Fantasies and
Reflections on Self and Soul*. London: Penguin Books

Holland, J. 1997. "Emergence,"

In *Philosophica*. 1997; 59(1): 11-40.

Holland, J. 1998. *Emergence: From Chaos to Order*. Massachusetts: Helix
Books.

Hume, D. 1969 [1739]: *A Treatise on Human Nature*. London: Penguin Books

IJsseling, S. 1997: *Mimesis. On Appearing and Being*. The Netherlands: Kok
Pharos Publishing House.

Juarrero, A. 1999. *Dynamics in Action. Intentional Behaviour as a Complex
System*. Cambridge, Massachusetts: MIT Press.

Juarrero, A. 2000. "Dynamics in Action: Intentional Behaviour as a Complex
System" *Emergence*, Volume #2, Issue #2. 2000.

Kant, I. (1990/1781): *Critique of Pure Reason*. New York: Prometheus Books

Kauffman, S.A. 1993. *The Origins of Order: Self-organization and Selection in Evolution*. New York: Oxford University Press.

Kauffman, S.A. 1995: *At Home in the Universe. The Search for the Laws of Complexity*. Harmondsworth: London.

Kenny, A 1989: *The Metaphysics of Mind*. Oxford: Oxford University Press.

Kuhn, T (1970) *The Structure of Scientific Revolutions*. Chicago: Chicago University Press.

Levin, J.D. 1992: *Theories of the Self*. Washington: Hemisphere Publishing Corporation

Lewin, R. 1993. *Complexity. Life on the Edge of Chaos*. London: Phoenix.

Lewin, R. and Regine, B. "Leading at the Edge: How Leaders Influence Complex Systems." *Emergence*, 2(2), 2000, pp.5-23.

Luria, A.R. (1998): "Mind and Brain: Luria's Philosophy"

in *The Oxford Companion to the Mind*. R.L. Gregory (ed.). Oxford: Oxford University Press.

Nagel, T. 1982: "What is it Like to be a Bat,"

in *The Mind's I. Fantasies and Reflections on Self and Soul*. Hofstadter, D.R. and Dennet, D.C. London: Penguin

Nathan, P. W. 1998: "Nervous System"

in *The Oxford Companion to the Mind*. R.L. Gregory (ed.). Oxford: Oxford University Press.

Padel, J. H. (1998): "Freudianism: Later Developments"

In *The Oxford Companion to the Mind*. R. L. Gregory (ed.). Oxford: Oxford University Press.

Peak, D and Frome, M. 1994: *Chaos Under Control. The Art and Science of Complexity*. New York: W. H. Freeman and Company.

Plato, 1981: *The Republic*. Harmondsworth: Penguin Classics

Popper, K. R., and Eccles, J.C. 1977: *The Self and its Brain*. Berlin: Springer-Verlag.

Prigogine, I. (1980). *From Being to Becoming. Time and Complexity in the Physical Sciences*. San Francisco: Freeman and Company.

Prigogine, I. and Stengers, I. 1984. *Order Out of Chaos: Man's New Dialogue with Nature*. London: Heinemann.

Radcliffe, E.S. 2000: *On Hume*. California: Wadsworth

Restak, R. 2001: *The Secret Life of the Brain*. The Dana Press and the Joseph Henry Press.

Rorty, R 1990: *Philosophy and the Mirror of Nature*. Oxford: Blackwell

Ryle, G. 1960: *The Concept of Mind*. London: Hutchinson

Searle, J. 1982: "Minds, Brains, Programmes"

in *The Mind's I. Fantasies and Reflections on Self and Soul*.

Hofstadter, D.R. and Dennet, D.C. London: Penguin

Sperry, R.W. 1998: "Consciousness and Causality"

In *The Oxford Companion to the Mind*. R.L. Gregory (ed.). Oxford:
Oxford University Press.

Sternberg, R.J. 1995: *In Search of the Human Mind*. Ford Worth: Harcourt
Brace College Publishers.

Strinati, D. 1995: *An Introduction to Theories of Popular Culture*. New
York: Routledge.

Solomon, R. C. and Higgins, K. M. 1996: *A Short History of Philosophy*.
Oxford: Oxford University Press.

Strachey, 1986 (ed.) in

Freud, S. [1950]: *Project for a Scientific Psychology*. Standard Edition,
volume 1, London: Hogarth Press and the Institute of Psycho-Analysis.

Toulmin, S. 1990: *Cosmopolis The Hidden Agenda of Modernity*, The Free
Press, New York.

Trevarthen, C. 1998: "Brain Development"

In: *The Oxford Companion to the Mind*. R. L. Gregory (ed.). Oxford:
Oxford University Press.

Verdenius, W. J. (1971): "Plato's Doctrine of Artistic Imitation"

in: Vlastos, G. (Ed): *Plato. A Collection of Critical Essays II*. New York: Doubleday and Company, Inc.

Von Bertalanffy, L. 1973: *General Systems Theory. Foundations, Developments, Applications*. Harmondsworth: Penguin.

Wuketits, F. M. 1998 "Emerging systems",

in: Koch, W. A. 1998: *Systems: new paradigms for the human sciences*. Berlin: De Gruyter.

Young, P. W. 1998: "Memory"

In *The Oxford Companion to the Mind*. R. L. Gregory (ed.). Oxford: Oxford University Press.

Zangwill, O. L. 1998a: "Sigmund Freud"

in *The Oxford Companion to the Mind*. R. L. Gregory (ed.). Oxford: Oxford University Press.

Zangwill, O. L. 1998b: "Freud on Mental Structure"

in *The Oxford Companion to the Mind*. R.L. Gregory (ed.). Oxford: Oxford University Press.